

Rec'd PCT 10/533 232
29 APR 2005 #2

PI 1142163

RECEIVED

15 MAR 2004

WIPO

PCT

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office

March 10, 2004

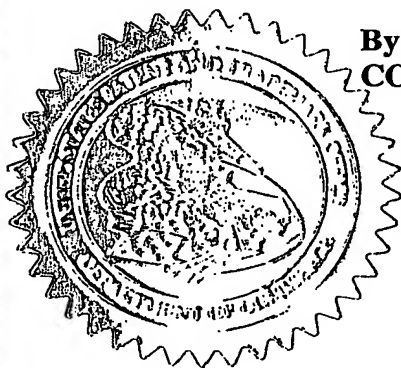
THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM
THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK
OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT
APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A
FILING DATE.

APPLICATION NUMBER: 60/436,565

FILING DATE: December 26, 2002

RELATED PCT APPLICATION NUMBER: PCT/US03/41200

BEST AVAILABLE COPY



By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

P. R. GRANT
Certifying Officer

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

12/26/02


Approved for use through 10/31/2002 OMB 055-E0032
 Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.
 U.S. Patent and Trademark Office, U.S. DEPARTMENT OF COMMERCE

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

Express Mail Label No.

EV03824713 US

INVENTOR(S)			
Given Name (first and middle) (if any)	Family Name or Surname	Residence (City and either State or Foreign Country)	
Brens	Coope	La Jolla CA	
<input type="checkbox"/> Additional inventors are being named on the <input type="checkbox"/> separately numbered sheets attached hereto.			
TITLE OF THE INVENTION (500 characters max)			
CELL PROLIFERATION-RELATED POLYPEPTIDES AND USES THEREFOR			
Direct all correspondence to: <input checked="" type="checkbox"/> Customer Number <input type="checkbox"/> Type Customer Number here			
<input checked="" type="checkbox"/> Customer Number 23483		 23483 PATENT TRADEMARK OFFICE	
OR			
<input type="checkbox"/> Firm or Individual Name		<input type="checkbox"/> Telephone Number	
Address		Address	
City		State	ZIP
Country	Telephone	Fax	
ENCLOSED APPLICATION PARTS (check all that apply)			
<input checked="" type="checkbox"/> Specification Number of Pages	237	<input type="checkbox"/> CD(s) Number	
<input checked="" type="checkbox"/> Drawing(s) Number of Sheets	109	<input type="checkbox"/> Other (specify)	
<input checked="" type="checkbox"/> Application Data Sheet See 37 CFR 1.76			
METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT			
<input type="checkbox"/> Applicant claims small entity status. See 37 CFR 1.27.		FILING FEE AMOUNT (\$)	
<input type="checkbox"/> A check or money order is enclosed to cover the filing fees.		160.00	
<input checked="" type="checkbox"/> The Commissioner is hereby authorized to charge filing fees on credit any overpayment to Deposit Account Number		08-0219	
<input type="checkbox"/> Payment by credit card Form PTO-2038 is attached.			
<input checked="" type="checkbox"/> The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.			
<input type="checkbox"/> No.			
<input type="checkbox"/> Yes the name of the U.S. Government agency and the Government contract number are			

Respectfully submitted

SIGNATURE

Date 12/26/2002

TYPED or PRINTED NAME Nancy Chiu, Ph.D.

REGISTRATION NO.

(if appropriate)

Docket Number

TELEPHONE (617) 526-6048

TMRI 025P 10984

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

This collection of information is required by 37 CFR 1.51. The information is used by the public to file (and by the PTO to process) a provisional application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 8 hours to complete, including gathering, preparing and submitting the complete provisional application to the PTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, Washington, D.C. 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Box Provisional Application Assistant, Commissioner for Patents, Washington, D.C. 20231.

EXPRESS MAIL LABEL NO. 310382713JHS
DATE OF DEPOSIT December 26, 2002

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number. Approved for use through 10/31/2002. OMB 0851-9031. U.S. Patent and Trademark Office, U.S. DEPARTMENT OF COMMERCE

FEE TRANSMITTAL for FY 2003

Patent fees are subject to annual revision.

☐ Applicant claims small entity status. See 37 CFR 1.27.

TOTAL AMOUNT OF PAYMENT (\$) 160.00

Complete if Known

Application Number: TBA
Filing Date: 12/26/2002
First Named Inventor: Bret Cooper
Examiner Name: TBA
Art Unit: TBA
Attorney Docket No.: FMRI-025P (109845, 154)

METHOD OF PAYMENT (check all that apply)

☒ Check ☒ Credit card ☐ Money Order ☐ Other ☐ None
☒ Deposit Account
Deposit Account Number: 08 0219
Deposit Account Name: Hale and Don LLP
The Commissioner is authorized to: (check all that apply)
☒ Charge fee(s) indicated below ☒ Credit any overpayments
☒ Charge any additional fee(s) during the pendency of this application
☐ Charge fee(s) indicated below, except for the filing fee to the above identified deposit account.

FEE CALCULATION

1. BASIC FILING FEE

Large Entity Fee Code (\$)	Small Entity Fee Code (\$)	Fee Description	Fee Paid
1001 1740	2001 370	Utility filing fee	
1002 330	2002 165	Design filing fee	
1003 510	2003 255	Plant filing fee	
1004 740	2004 370	Reissue filing fee	
1005 160	2005 80	Provisional filing fee	160.00

SUBTOTAL (1) (\$) 160.00

2. EXTRA CLAIM FEES FOR UTILITY AND REISSUE

Total Claims	Independent Claims	Multiple Dependent	Extra Claims	Fee from below	Fee Paid
20	3			0.00	0.00

Large Entity Fee Code (\$)	Small Entity Fee Code (\$)	Fee Description
1202 18	2202 19	Claims in excess of 20
1201 84	2201 42	Independent claims in excess of 3
1203 280	2203 140	Multiple dependent claim, if not paid over original patent
1204 280	2204 140	Reissue independent claim
1205 18	2205 19	Reissue claims in excess of 20 and over original patent

SUBTOTAL (2) (\$) 0.00

or number previously paid, if greater. For Reissues, see above.

FEE CALCULATION (continued)

3. ADDITIONAL FEES

Large Entity Fee Code (\$)	Small Entity Fee Code (\$)	Fee Description	Fee Paid
1051 130	2051 65	Surcharge late filing fee or oath	
1052 50	2052 25	Surcharge late provisional filing fee or cover sheet	
1053 130	2053 65	Non-English specification	
1812 2520	2812 2520	For filing a request for ex parte reexamination	
1804 920	2804 920	Requesting publication of SIR prior to Examiner action	
1805 1840	2805 1840	Requesting publication of SIR after Examiner action	
2251 210	2251 105	Extension to reply within first month	
2252 400	2252 200	Extension to reply within second month	
2253 920	2253 460	Extension to reply within third month	
2254 1440	2254 720	Extension to reply within fourth month	
2255 1960	2255 980	Extension to reply within fifth month	
2401 320	2401 160	Notice of Appeal	
2402 320	2402 160	Filing a brief in support of an appeal	
2403 280	2403 140	Request for oral hearing	
2451 510	2451 255	Petition to institute a public use proceeding	
2452 110	2452 55	Petition to revive unavoidable	
2453 280	2453 140	Petition to revive unintentional	
2501 280	2501 140	Utility issue fee (or reissue)	
2502 60	2502 30	Design issue fee	
2503 310	2503 155	Plant issue fee	
2460 130	2460 65	Petitions to the Commissioner	
1807 50	2807 25	Processing fee under 37 CFR 1.17(d)	
1806 180	2806 90	Submission of information Disclosure Statement	
8021 40	28021 20	Recording each patent assignment property (times number of properties)	
1809 740	2809 370	Filing a submission after final rejection (37 CFR 1.129(a))	
1810 740	2810 370	For each additional invention to be examined (37 CFR 1.129(b))	
1801 740	2801 370	Request for Continued Examination (RCE)	
1802 900	2802 450	Request for expedited examination of a design application	

Other fee (specify)

Reduced by Basic Filing Fee Paid

SUBTOTAL (3) (\$) 0.00

SUBMITTED BY

Name (Print/Type): Nancy Chau, Ph.D.
Signature: [Signature]
Registration No. (Attorney/Agent): 431545
Telephone: (617) 526-6048
Date: 12/26/2002

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

This collection of information is required by 37 CFR 1.17 and 1.27. The information is required to obtain or retain a benefit by the public which is to file (and by USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete the amount of time you require to complete this form and/or suggestions for reducing this burden should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: [Address]

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Attorney Docket No. 109845.154

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s): Bret Cooper

Serial No.: TBA

Filed: December 26, 2002

Title: CELL PROLIFERATION-RELATED POLYPEPTIDES
AND USES THEREFOR

Box Provisional Application
Assistant Commissioner for Patents
Washington, D.C. 20231

TRANSMITTAL LETTER

Dear Sir:

Enclosed herewith for filing in the above-referenced provisional patent application are the following documents:

1. United States provisional patent application entitled:

**CELL PROLIFERATION-RELATED POLYPEPTIDES
AND USES THEREFOR**

and naming as the inventor(s):

Bret Cooper

The provisional application comprising:

237 pages of Specification

109 sheets of Figures

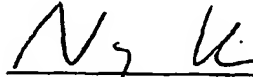
2. Provisional Application For Patent Cover Sheet.
3. Application Data Sheet.
4. Fee Transmittal.
5. Return Receipt Postcard.

Attorney Docket No. 109845.154

PATENT

The Commissioner is authorized to debit our Deposit Account No. 08-0219 for the filing fee for a provisional application and any other fee deemed necessary in connection with the filing of the above-identified provisional application.

Respectfully submitted,



Nancy Chiu, Ph.D.

Reg. No. 43,545

Attorney for Applicant

Date: December 26, 2002

Hale and Dorr LLP
60 State Street
Boston, MA 02109
Telephone: 617-526-6048
Facsimile: 617-526-5000

EXPRESS MAIL LABEL NO. EV038247131US
DATE OF DEPOSIT December 26, 2002

Application Data Sheet

Application Information

Application number::
Filing Date:: December 26, 2002
Application Type:: Regular
Subject Matter:: Provisional
Suggested Classification::
Suggested Group Art::
CD-ROM or CD-R?: None
Computer Readable Form (CRF)?:: No
Title:: CELL PROLIFERATION-RELATED
POLYPEPTIDES AND USES THEREFOR
Attorney Docket Number:: 109845.154
Request for Early Publication?:: No
Request for Non-Publication?:: No
Suggested Drawing Figure::
Total Drawing Sheets:: 109
Small Entity?:: No
Licensed US Govt. Agency:: No
Contract or Grant Numbers::
Secrecy Order in Parent Appl.?:: No

Applicant Information

Applicant Authority Type:: Inventor
Primary Citizenship Country:: US
Status:: Full Capacity
Given Name:: Bret
Middle Name::
Family Name:: Cooper
Name Suffix::
City of Residence:: La Jolla

EXPRESS MAIL LABEL NO. EV038247131US
DATE OF DEPOSIT December 26, 2002

State or Province of Residence:: CA
 Country of Residence:: US
 Street of Mailing Address:: 3314 Via Alicante
 City of Mailing Address:: La Jolla
 State or Province of Mailing Address:: CA
 Country of Mailing Address:: US
 Postal or Zip Code of Mailing Address:: 93037

Correspondence Information

Correspondence Customer Number:: 23483
 Phone Number:: 617-526-6048
 Fax Number:: 617-526-5000
 E-Mail Address:: nancy.chiu@haledorr.com

Representative Information

Representative Customer Number:: 23483

Domestic Priority Information

Application::	Continuity Type::	Parent Application:: [20 character max]	Parent Filing Date:: [8 characters mm/dd/yy]

Foreign Priority Information

Country:: [50 character max]	Application Number:: [20 character max]	Filing Date:: [8 characters mm/dd/yy]	Priority Claimed:: Yes or No

EXPRESS MAIL LABEL NO. EV038247131US
DATE OF DEPOSIT December 26, 2002

Assignee Information

Assignee Name::	Torrey Mesa Research Institute
Street of Mailing Address::	3115 Merryfield Row
City of Mailing Address::	San Diego
State or Province of Mailing Address::	CA
Country of Mailing Address::	US
Postal or Zip Code of Mailing Address::	92121-1125

CELL PROLIFERATION-RELATED POLYPEPTIDES AND USES THEREFOR**Background of the Invention**

This invention relates to the field of transgenic plants.

As some of the major human staples, monocot plants such as rice, corn, and wheat
5 have been a target of genetic engineering for higher yields and resistance to diseases,
pests, and environmental stresses of various kinds. The timing of the transition from
vegetative growth to flowering, for example, is an important step in plant development
that determines the quality and quantity of most crop species by affecting the balance
between vegetative and reproductive growth. Therefore, control of flowering time in
10 genetically engineered cereal crops is important in agriculture. Knowledge of the
proteins and molecular interactions associated with cell cycle processes, development,
and stress response in monocot plants, such as rice, could lead to important applications
in agriculture. Modulation of these interactions may be exploited to effect changes in
plant development or growth that would result in increased crop yield and, in addition,
15 may be used to increase tolerance to environmental stress conditions.

Similarly, the development of plant organs (*e.g.*, root and stem), and the ability of
a plant to respond to stress and to defend itself from insects and pathogens are likewise
important targets for genetic engineering. Genes encoding proteins involved in the plant
response to pathogens are important to agriculture, as their discovery may allow genetic
20 manipulation of crops to obtain plants with enhanced or reduced disease resistance.

Thus, there is a need to identify proteins that are involved in plant growth
(including cell cycle and senescence), plant development, and plant responses to stress.
Knowledge of the interactions of such proteins will allow opportunities to produce
enhanced food crops.
25

Summary of the Invention

The invention provides proteins and nucleic acid molecules encoding such
proteins that are involved in the control and regulation of plant maturation and
development, including proliferation, senescence, disease-resistance, stress-resistance,
30 and differentiation. The invention provides compositions comprising at least one of the

proteins described herein, as well as methods for using the proteins disclosed herein to affect plant maturation, development, and responses to stress.

In one aspect, the invention provides an isolated nucleic acid molecule encoding a cell proliferation-related polypeptide, wherein the polypeptide binds to a fragment of a protein selected from the group consisting of OsE2F1, Os018989-4003, OsE2F2, OsS49462, OsCYCOS2, OsMADS45, OsRAP1B, OsMADS6, OsFDRMADS8, OsMADS3, OsMADS5, OsMADS15, OsHOS59, OsGF14-c, OsDAD1, Os006819-2510, OsCRTC, OsSGT1, OsERP, OsCHIB1, OsCS, OsPP2A-2, and OsCAA90866. In certain embodiments, the isolated nucleic acid molecule is derived from rice (*i.e.*, *Oryza sativa*).
In certain embodiments, the invention provides an isolated nucleic acid molecule comprising a nucleotide sequence substantially similar to the nucleotide sequence of the nucleic acid molecule encoding a cell proliferation-related polypeptide of the invention.

In certain embodiments, the protein consists of an amino acid sequence selected from a sequence shown in Figure 7. In some embodiments, the nucleic acid molecule comprises or consists of the nucleic acid sequence of a sequence selected from the group consisting of a sequence shown in Figure 8, Figure 9, Figure 10, Figure 11, Figure 12, Figure 13, Figure 14, Figure 15, or Figure 16.

In certain embodiments, a cell introduced with a nucleic acid molecule of the invention has a different cell proliferation rate as compared to a cell not introduced with the nucleic acid molecule.

In another aspect, the invention features a polypeptide encoded by the nucleic acid molecule of the invention.

In yet another aspect, the invention features an isolated cell proliferation-related polypeptide, wherein the polypeptide binds to a fragment of a protein selected from the group consisting of OsE2F1, Os018989-4003, OsE2F2, OsS49462, OsCYCOS2, OsMADS45, OsRAP1B, OsMADS6, OsFDRMADS8, OsMADS3, OsMADS5, OsMADS15, OsHOS59, OsGF14-c, OsDAD1, Os006819-2510, OsCRTC, OsSGT1, OsPN31085, OsCHIB1, OsCS, OsPP2A-2, and OsCAA90866. In some embodiments, the invention features an isolated polypeptide comprising or consisting of an amino acid sequence substantially similar to the amino acid sequence of an isolated cell proliferation-related polypeptide of the invention.

In yet another aspect, the invention features an expression cassette comprising a nucleic acid molecule encoding a cell proliferation-related polypeptide of the invention. In certain embodiments, the expression cassette further comprises a regulatory element such that the cell proliferation-related polypeptide is expressed by a host cell comprising the expression cassette. In certain embodiments, the invention features a host cell comprising the expression cassette. In further embodiments, the invention features a transgenic plant comprising the expression cassette.

In another aspect, the invention features a method for modulating the proliferation of a plant cell comprising introducing an isolated nucleic acid molecule encoding a cell proliferation-related polypeptide into the plant cell, wherein the polypeptide binds to a fragment of a protein selected from the group consisting of OsE2F1, Os018989-4003, OsE2F2, OsS49462, OsCYCOS2, OsMADS45, OsRAP1B, OsMADS6, OsFDRMADS8, OsMADS3, OsMADS5, OsMADS15, OsHOS59, OsGF14-c, OsDAD1, Os006819-2510, OsCRTC, OsSGT1, OsERP, OsCHIB1, OsCS, OsPP2A-2, and OsCAA90866, wherein the polypeptide is expressed by the cell.

In yet another aspect, the invention features a method for modulating the proliferation of a plant cell comprising introducing an isolated nucleic acid molecule encoding a cell proliferation-related polypeptide into the plant cell, wherein the polypeptide binds to a fragment of a protein selected from the group consisting of OsE2F1, Os018989-4003, OsE2F2, OsS49462, OsCYCOS2, OsMADS45, OsRAP1B, OsMADS6, OsFDRMADS8, OsMADS3, OsMADS5, OsMADS15, OsHOS59, OsGF14-c, OsDAD1, Os006819-2510, OsCRTC, OsSGT1, OsERP, OsCHIB1, OsCS, OsPP2A-2, and OsCAA90866, wherein the polypeptide is expressed by the cell.

In another aspect, the invention features a method for modulating the proliferation of a plant cell comprising introducing an isolated nucleic acid molecule encoding a cell proliferation-related polypeptide into the plant cell, wherein the polypeptide binds to a fragment of a protein selected from the group consisting of OsE2F1, Os018989-4003, OsE2F2, OsS49462, OsCYCOS2, OsMADS45, OsRAP1B, OsMADS6, OsFDRMADS8, OsMADS3, OsMADS5, OsMADS15, OsHOS59, OsGF14-c, OsDAD1, Os006819-2510, OsCRTC, OsSGT1, OsERP, OsCHIB1, OsCS, OsPP2A-2, and OsCAA90866, wherein expression of the polypeptide encoded by the nucleic acid molecule is reduced in the cell.

Brief Description of the Drawings

Figure 1 is a schematic representation of the interactions between various, non-limiting, cell proliferation-related proteins of the invention. Arrows indicate interaction direction between DNA binding domain fused proteins (thick lined boxes or ovals) and activation domain fused proteins. Dotted boxes indicate previously published interactions. Ovals rather than boxes indicate that a protein fused to the DNA binding domain did not interact with other proteins. Circular arrows depict self-interactions. Dotted lines indicate amino acid similarity between proteins. Box colors denote functional classification: Purple, cell cycle; blue, development; rose, biotic stress; orange, abiotic stress; green, chloroplast; black, undefined role. (Note that included herewith is a small, color version of this figure, as well as a larger, black and white version of this figure).

Figure 2 is a schematic representation of the interactions between various, non-limiting, cell proliferation-related proteins of the invention. Arrows indicate interaction direction between DNA binding domain fused proteins (thick lined boxes or ovals) and activation domain fused proteins. Dotted boxes indicate previously published interactions. Ovals rather than boxes indicate that a protein fused to the DNA binding domain did not interact with other proteins. Circular arrows depict self-interactions. Dotted lines indicate amino acid similarity between proteins. Box colors denote functional classification: Purple, cell cycle; blue, development; rose, biotic stress; orange, abiotic stress; green, chloroplast; black, undefined role.

Figure 3A is a schematic representation showing an amino acid alignment of various, non-limiting, cell proliferation-related proteins of the invention.

Figure 3B is a schematic representation showing a phylogenetic tree of the proteins whose amino acid sequences are aligned in Figure 3A.

Figure 4 is a schematic representation of the interactions between various, non-limiting, cell proliferation-related proteins of the invention. Arrows indicate interaction direction between DNA binding domain fused proteins (thick lined boxes or ovals) and activation domain fused proteins. Dotted boxes indicate previously published interactions. Ovals rather than boxes indicate that a protein fused to the DNA binding domain did not interact with other proteins. Circular arrows depict self-interactions.

Dotted lines indicate amino acid similarity between proteins. Box colors denote functional classification: Purple, cell cycle; blue, development; rose, biotic stress; orange, abiotic stress; green, chloroplast; black, undefined role.

5 Figure 5 is a schematic representation of the interactions between various, non-limiting, cell proliferation-related proteins of the invention. Arrows indicate interaction direction between DNA binding domain fused proteins (thick lined boxes or ovals) and activation domain fused proteins. Dotted boxes indicate previously published interactions. Ovals rather than boxes indicate that a protein fused to the DNA binding domain did not interact with other proteins. Circular arrows depict self-interactions.

10 Dotted lines indicate amino acid similarity between proteins. Box colors denote functional classification: Purple, cell cycle; blue, development; rose, biotic stress; orange, abiotic stress; green, chloroplast; black, undefined role.

15 Figure 6 is a schematic representation of the interactions between various, non-limiting, cell proliferation-related proteins of the invention. Arrows indicate interaction direction between DNA binding domain fused proteins (thick lined boxes or ovals) and activation domain fused proteins. Dotted boxes indicate previously published interactions. Ovals rather than boxes indicate that a protein fused to the DNA binding domain did not interact with other proteins. Circular arrows depict self-interactions.

20 Dotted lines indicate amino acid similarity between proteins. Box colors denote functional classification: Purple, cell cycle; blue, development; rose, biotic stress; orange, abiotic stress; green, chloroplast; black, undefined role.

Figure 7 lists the amino acid sequences of the target proteins (*i.e.*, "bait" proteins) used in the present invention in single letter code (with the amino-terminal M being amino acid no. 1, and the * being the stop signal).

25 Figure 8 lists the nucleotide sequences encoding the proteins identified in Example I.

Figure 9 lists the nucleotide sequences encoding the proteins identified in Example II.

30 Figure 10 lists the nucleotide sequences encoding the proteins identified in Example III.

Figure 11 lists the nucleotide sequences encoding the proteins identified in Example IV.

Figure 12 lists the nucleotide sequences encoding the proteins identified in Example V.

5 Figure 13 lists the nucleotide sequences encoding the proteins identified in Example VI.

Figure 14 lists the nucleotide sequences encoding the proteins identified in Example VII.

10 Figure 15 lists the nucleotide sequences encoding the proteins identified in Example VIII.

Figure 16 lists the nucleotide sequences encoding the proteins identified in Example IX.

15

Detailed Description of the Preferred Embodiments

All of the patents (including published patent applications) and publications (including GenBank sequence references), which are cited herein reflect the knowledge in the art and are hereby incorporated by reference in entirety to the same extent as if each were specifically stated to be incorporated by reference. Any inconsistency between these patents and publications and the present disclosure shall be resolved in favor of the present disclosure.

This invention stems from the recognition that proteins that participate in plant cell proliferation, including those proteins involved in cell cycle regulation, plant development, stress response (both biotic and abiotic), and senescence, may be targets for genetic manipulation or for compounds that modify their level or activity, thereby modulating the proliferation of the plant cell. The identification of genes encoding these proteins in rice allows for the development of methods for controlling plant growth and proliferation. For example, methods for controlling cell proliferation and differentiation can facilitate or retard plant development and promote regeneration. Similarly, methods for controlling stress response can facilitate plant responses and endurance to abiotic stress (*e.g.*, temperature or salinity) or biotic stress (*e.g.*, pathogen infection). Such methods may involve the application of compounds to crops or the engineering of plants in which the level and/or activity of a plant cell proliferation protein is modulated for a time and under conditions sufficient to modify or control cell proliferation.

In one aspect, the invention provides an isolated nucleic acid molecule encoding a cell proliferation-related polypeptide, wherein the polypeptide binds to a fragment of a protein selected from the group consisting of OsE2F1, Os018989-4003, OsE2F2, OsS49462, OsCYCOS2, OsMADS45, OsRAP1B, OsMADS6, OsFDRMADS8, OsMADS3, OsMADS5, OsMADS15, OsHOS59, OsGF14-c, OsDAD1, Os006819-2510, OsCRTC, OsSGT1, OsERP, OsCHIB1, OsCS, OsPP2A-2, and OsCAA90866. In certain embodiments, the isolated nucleic acid molecule is derived from rice (*i.e.*, *Oryza sativa*).

The invention encompasses isolated nucleic acid molecule or protein (or polypeptide) compositions. As used herein, an "isolated" nucleic acid molecule or an "isolated" protein is a nucleic acid molecule or a protein, respectively, that is substantially free from components that normally accompany or interact with it in nature.

It should be noted that a nucleic acid molecule or protein is isolated as used in accordance with the invention even if it is not alone but, rather, surrounded by other molecules, so long as those molecules are not molecules which normally accompany or interact with the isolated nucleic acid molecule or protein in nature. For example, a gene from a wheat cell is isolated if it is expressed in a non-wheat plant cell (*e.g.*, a rice cell).

An "isolated" or "purified" nucleic acid molecule or protein, or biologically active portion thereof, is substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. In certain embodiments, an "isolated" nucleic acid is free of sequences (*e.g.*, protein encoding sequences) that naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb, or 0.1 kb of nucleotide sequences that naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. A protein that is substantially free of cellular material includes preparations of protein or polypeptide having less than about 30%, 20%, 10%, or 5%, (by dry weight) of contaminating protein. When the protein of the invention, or biologically active portion thereof, is recombinantly produced, culture medium represents less than about 30%, 20%, 10%, or 5% (by dry weight) of chemical precursors or non-protein of interest chemicals.

As used herein, by a "cell proliferation-related polypeptide", is meant a protein or polypeptide (note that these two terms are used interchangeably throughout) that is involved with cell proliferation, particularly plant cell proliferation. Such a polypeptide may be involved in the increase in cell proliferation; conversely, such a polypeptide may be involved in the abrogation of cell proliferation. Moreover, the polypeptide may be involved in cell proliferation only, for example, when the cell is exposed to a stress (*e.g.*, biotic or abiotic). In addition, the polypeptide may be involved in cell proliferation only when the cell is differentiating or developing. A "cell proliferation-related polypeptide" of the invention is identified by the ability of an increase or decrease in the level of expression of such a polypeptide in a cell to modulate the rate of that cell's proliferation,

whether alone or together with some other stimuli (*e.g.*, presence of growth factor, presence of stress).

As used herein, by the term, "binds" means that a cell proliferation-related polypeptide preferentially interacts with a stated target molecule. In some embodiments, that interaction allows a biological read-out (*e.g.*, a positive in the yeast two-hybrid system). In some embodiments, that interaction is measurable (*e.g.*, a K_D of at least 10^{-5} M).

The present inventors have isolated, cloned and characterized rice (*O. sativa*)-derived cDNAs encoding plant proteins that interact with OsE2F1, Os018989-4003, OsE2F2, OsS49462, OsCYCOS2, OsMADS45, OsRAP1B, OsMADS6, OsFDRMADS8, OsMADS3, OsMADS5, OsMADS15, OsHOS59, OsGF14-c, OsDAD1, Os006819-2510, OsCRTC, OsSGT1, OsERP, OsCHIB1, OsCS, OsPP2A-2, and OsCAA90866 in the yeast two-hybrid system.

The yeast two-hybrid system is a well known system which is based on the finding that most eukaryotic transcription activators are modular (see, *e.g.*, Gyuris *et al.*, *Cell* 1993, 75: 791-803, 1993; Feys *et al.*, *EMBO J.* 20: 5400-5411, 2001 The Yeast Two-Hybrid System, Bartel and Fields (eds.), Oxford Press, 1997. All yeast two-hybrid systems use 1) a plasmid that directs the synthesis of a "bait" (a known protein which is brought to the yeast's DNA by being fused to a DNA binding domain), 2) one or more reporter genes ("reporters") with upstream binding sites for the bait, and 3) a plasmid that directs the synthesis of proteins fused to activation domains and other useful moieties ("activation tagged proteins", or "prey").

In all of the Examples described below, an automated, high-throughput yeast two-hybrid assay technology (provided by Myriad Genetics Inc., Salt Lake City, UT) was used to search for protein interactions with the bait proteins. Briefly, the target protein (*e.g.*, OsE2F1) was expressed in yeast as a fusion to the DNA-binding domain of the yeast Gal4p. DNA encoding the target protein or a fragment of this protein was amplified from cDNA by PCR or prepared from an available clone. The resulting DNA fragment was cloned by ligation or recombination into a DNA-binding domain vector (*e.g.*, pGBT9, pGBT.C, pAS2-1) such that an in-frame fusion between the Gal4p and

target protein sequences was created. The resulting construct, the target gene construct, was introduced by transformation into a haploid yeast strain.

5 A screening protocol was then used to search the individual baits against two activation domain libraries of greater than five million cDNA clones of assorted peptide motifs. The libraries were derived from RNA isolated from leaves, stems, and roots of rice plants grown in normal conditions plus tissues from plants exposed to various stresses (input trait library), and from various seed stages, callus, and early and late panicle (output trait library). To screen, a library of activation domain fusions (*i.e.*, *O. sativa* cDNA cloned into an activation domain vector) was introduced by transformation
10 into a haploid yeast strain of the opposite mating type. The yeast strain that carried the activation domain constructs contained one or more Gal4p-responsive reporter gene(s), whose expression can be monitored. Non-limiting examples of some yeast reporter strains include Y190, PJ69, and CBY14a.

Yeast carrying the target gene construct was combined with yeast carrying the
15 activation domain library. The two yeast strains mated to form diploid yeast and were plated on media that selected for expression of one or more Gal4p-responsive reporter genes. Thus, both hybrid proteins (*i.e.*, the target "bait" protein and the activation domain "prey" protein) were expressed in a yeast reporter strain where an interaction between the test proteins results in transcription of the reporter genes *TRP1* and *LEU2*, allowing
20 growth on selective medium lacking tryptophan and leucine. Colonies that arose after incubation were selected for further characterization. The activation domain plasmid was isolated from each colony obtained in the two-hybrid search. The sequence of the insert in this construct was obtained by the sequence analysis (*e.g.*, Sanger's dideoxy nucleotide chain termination method—see Ausubel *et al.*, Current Protocols in Molecular Biology,
25 John Wiley & Sons, New York, NY 1988, including updates up to 2002). Thus, the identity of positives obtained from these searches was determined by sequence analysis against proprietary and public (*e.g.*, GenBank) nucleic acid and protein databases.

Interaction of the activation domain fusion with the target protein was confirmed by testing for the specificity of the interaction. The activation domain construct was co-
30 transformed into a yeast reporter strain with either the original target protein construct or a variety of other DNA-binding domain constructs. Expression of the reporter genes in

the presence of the target protein but not with other test proteins indicated that the interaction was genuine.

To further characterize the genes encoding the interacting proteins, the gene sequences of the baits and preys were compared with the gene fragments represented on
5 TMRI's proprietary GeneChip® Rice Genome Array (Affymetrix, Santa Clara, CA) (see Zhu *et al.*, *Plant Physiol. Biochem.* 39: 221-242, 2001). The rice genome array contained 25-mer oligonucleotide probes with sequences corresponding to the 3' ends of 21,000 predicted open reading frames found in approximately 42,000 contigs that make up the rice genome map (see Goff *et al.*, *Science* 296: 92-100, 2002). Sixteen different probes
10 were used to measure the expression level of each gene. The sequences of the probes http://tmri.org/gene_exp_web/). The expression value was determined based on the expression level minus the noise background associated with each probe. Experiments included evaluating the differential gene expression from various plant tissues comprising seed, root, leaf and stem, panicle, and pollen. Gene expression was also measured in
15 plants exposed to environmental cold (*i.e.*, 14°C), osmotic pressure (growth media supplemented with 260 mM mannitol), drought (media supplemented with 25% polyethylene glycol 8000), salt (media supplemented with 150 mM NaCl), ABA-inducible stresses (media supplemented with 50 uM ABA; see Chen *et al.*, *Plant Cell* 14: 559-574, 2002), infection by the fungal pathogen *Magnaporthe grisea*, and treatment
20 with plant hormones (jasmonic acid (JA, 100 uM), gibberellin (GA3, 50 uM), and abscisic acid) and with herbicides (BAP (10 uM), 2,4-D, and BL2 (10 uM)).

In certain embodiments, the invention provides an isolated nucleic acid molecule comprising a nucleotide sequence substantially similar to the nucleotide sequence of the nucleic acid molecule encoding a cell proliferation-related polypeptide of the invention.

25 In a broad sense, the term "substantially similar", when used herein with respect to a nucleotide sequence, means a nucleotide sequence corresponding to a reference nucleotide sequence (*i.e.*, a nucleotide sequence of a nucleic acid molecule encoding a cell proliferation-related protein of the invention), wherein the corresponding sequence encodes a polypeptide having substantially the same structure as the polypeptide encoded
30 by the reference nucleotide sequence. In some embodiments, the substantially similar nucleotide sequence encodes the polypeptide encoded by the reference nucleotide

sequence (*i.e.*, although the nucleotide sequence is different, the encoded protein has the same amino acid sequence). In some embodiments, "substantially similar" refers to nucleotide sequences having at least 50% sequence identity, or at least 60%, 70%, 80% or 85%, or at least 90% or 95%, or at least 96%, 97% or 99% sequence identity compared to a reference sequence containing nucleotide sequences encoding one of the cell proliferation-related proteins of the invention (e.g., the proteins described below in the example).

Methods of alignment of sequences for comparison are well known in the art. Thus, the determination of percent identity between any two sequences can be accomplished using a mathematical algorithm. Non-limiting examples of such mathematical algorithms are the algorithm of Myers and Miller, *CABIOS* 4:11, 1988; the local homology algorithm of Smith *et al.*, *Adv. Appl. Math.* 2: 482, 1981; the homology alignment algorithm of Needleman and Wunsch *J. Mol. Biol.* 48:443, 1970; the search-for-similarity-method of Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444, 1988; the algorithm of Karlin and Altschul, *Proc. Natl. Acad. Sci. USA* 87:2264, 1990, modified as in Karlin and Altschul, *Proc. Natl. Acad. Sci. USA* 90:5873, 1993.

Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity. Such implementations include, but are not limited to: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, California); the ALIGN program (Version 2.0) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Version 8 (available from Genetics Computer Group (GCG), 575 Science Drive, Madison, Wisconsin, USA). Alignments using these programs can be performed using the default parameters. The CLUSTAL program is well described by Higgins *et al.*, *Gene* 73:237 1988; Higgins *et al.*, *CABIOS* 5:151, 1989; Corpet *et al.*, *Nuc. Acids Res.* 16:10881, 1988; Huang *et al.*, *CABIOS* 8:155, 1992; and Pearson *et al.*, *Meth. Mol. Biol.* 24:307, 1994. The ALIGN program is based on the algorithm of Myers and Miller, *supra*. The BLAST programs of Altschul *et al.*, *J. Mol. Biol.* 215:403, 1990, are based on the algorithm of Karlin and Altschul *supra*.

Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This

algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *J. Mol. Biol.*, 215:403, 1990). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when the cumulative alignment score falls off by the quantity X from its maximum achieved value, the cumulative score goes to zero or below due to the accumulation of one or more negative-scoring residue alignments, or the end of either sequence is reached.

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul (*Proc. Natl. Acad. Sci. USA*, 90:5873, 1993). One measure of similarity provided by the BLAST algorithm is the smallest sum probability ($P(N)$), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a test nucleic acid sequence is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid sequence to the reference nucleic acid sequence is less than about 0.1, or less than about 0.01, or less than about 0.001.

To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized as described in Altschul *et al.*, *Nuc. Acids Res.*, 25:3389, 1997. Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. See Altschul *et al.*, *supra*. When utilizing BLAST, Gapped BLAST, PSI-BLAST, the default parameters of the respective programs (e.g. BLASTN for nucleotide sequences, BLASTX for proteins) can be used. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength

(W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA*, 89:10915, 1989). (See
5 <http://www.ncbi.nlm.nih.gov>). Alignment may also be performed manually by inspection.

For purposes of the present invention, comparison of nucleotide sequences for determination of percent sequence identity to the sequences disclosed herein is made using the BlastN program (version 1.4.7 or later) with its default parameters or any
10 equivalent program. By "equivalent program" is intended any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide or amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by the BlastN program.

"Substantially similar" also refers to nucleotide sequences having at least 50%
15 identity, or at least 80% identity, or at least 95% identity, or at least 99% identity, to a region of nucleotide sequence encoding a BIOPATH protein and/or an FPD, wherein the nucleotide sequence comparisons are conducted using GAP analysis as described herein. The term "substantially similar" is specifically intended to include nucleotide sequences wherein the sequence has been modified to optimize expression in particular cells.

20 A polynucleotide including a nucleotide sequence "substantially similar" to the reference nucleotide sequence hybridizes to a polynucleotide including the reference nucleotide sequence in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 2X SSC, 0.1% SDS at 50°C, more desirably in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 1X SSC, 0.1% SDS
25 at 50°C, more desirably still in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.5X SSC, 0.1% SDS at 50°C, or in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 50°C, or in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 65°C.

30 The term "substantially similar", when used herein with respect to a protein or polypeptide, means a protein or polypeptide corresponding to a reference protein (*i.e.*, a

cell proliferation-related protein of the invention), wherein the protein has substantially the same structure and function as the reference protein, where only changes in amino acids sequence that do not materially affect the polypeptide function occur. When used for a protein or an amino acid sequence the percentage of identity between the

5 substantially similar and the reference protein or amino acid sequence is at least 30%, or at least 40%, 50%, 60%, 70%, 80%, 85%, or 90%, or at least 95%, or at least 99% with every individual number falling within this range of at least 30% to at least 99% also being part of the invention, using default GAP analysis parameters with the University of Wisconsin GCG (version 10), SEQWEB application of GAP, based on the algorithm of

10 Needleman and Wunsch, *J. Mol. Biol.* 48:443, 1970.

All of the cell proliferation-related proteins of the invention are related, and many interact with one another. Figures 1-6 are schematic representations showing the interrelatedness of the different cell proliferation-related proteins of the invention.

In certain embodiments, a target protein of the invention comprises or consists of

15 an amino acid sequence selected from a sequence shown in Figure 7. In some embodiments, the nucleic acid molecule comprises or consists of the nucleic acid sequence of a sequence selected from the group consisting of a sequence shown in Figure 8, Figure 9, Figure 10, Figure 11, Figure 12, Figure 13, Figure 14, Figure 15, or Figure 16.

20 In another aspect, the invention features a cell proliferation-related polypeptide encoded by the nucleic acid molecule of the invention. In certain embodiments, the cell proliferation-related polypeptide is isolated.

For example, a nucleic acid molecule of the invention can be introduced, under conditions for expression, into a host cell such that the host cell transcribes and translates

25 the nucleic acid molecule to produce a cell proliferation-related polypeptide. By "under conditions for expression" is meant that a nucleic acid molecule is positioned in the cell such that it will be expressed in that cell. For example, a nucleic acid molecule may be located downstream of a promoter that is active in the cell, such that the promoter will drive the expression of the polypeptide encoded for by the nucleic acid molecule in the

30 cell. Any regulatory sequence (*e.g.*, promoter, enhancer, inducible promoter) can be linked to the nucleic acid molecule; alternatively, the nucleic acid molecule may include

its own regulatory sequence such that it will be expressed (*i.e.*, transcribed and/or translated) in a cell.

Where the nucleic acid molecule of the invention is introduced into a cell under conditions of expression, that nucleic acid molecule can be said to be included in an expression cassette. Thus, the invention further provides a host cell comprising
5 expression cassette comprising a nucleic acid molecule encoding a cell proliferation-related polypeptide of the invention. Such an expression cassette includes, in addition to the nucleic acid molecule encoding a cell proliferation-related polypeptide of the invention, at least one regulatory sequence (*e.g.*, a promoter or enhancer).

10 In one non-limiting example, a plant promoter fragment may be employed which will direct expression of the gene in all tissue of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation
15 region, the 1'- or 2'-promoter derived from T-DNA of *Agrobacterium tumefaciens*, and other transcription initiation regions from various plant genes known to those of skill. Such genes include for example, the AP2 gene, ACT11 from *Arabidopsis* (Huang *et al.*, *Plant Mol. Biol.* 33:125, 1996), Cat3 from *Arabidopsis* (GenBank Accession No. U43147, Zhong *et al.*, *Mol. Gen. Genet.* 251:196, 1996), the gene encoding stearyl-acyl
20 carrier protein desaturase from *Brassica napus* (Genbank Accession No. X74782, Solocombe *et al.*, *Plant Physiol.* 104:1167, 1994), GPc1 from maize (GenBank Accession No. X15596, Martinez *et al.*, *J. Mol. Biol.* 208:551, 1989), and Gpc2 from maize (GenBank Accession No. U45855, Manjunath *et al.*, *Plant Mol. Biol.* 33:97, 1997).

Alternatively, the plant promoter may direct expression of the nucleic acid
25 molecules of the invention in a specific tissue or may be otherwise under more precise environmental or developmental control. Examples of environmental conditions that may effect transcription by inducible promoters include anaerobic conditions, elevated temperature, or the presence of light. Such promoters are referred to here as "inducible" or "tissue-specific" promoters. One of skill will recognize that a tissue-specific promoter
30 may drive expression of operably linked sequences in tissues other than the target tissue.

Thus, as used herein a tissue-specific promoter is one that drives expression preferentially in the target tissue, but may also lead to some expression in other tissues as well.

Examples of promoters under developmental control include promoters that initiate transcription only (or primarily only) in certain tissues, such as fruit, seeds, or flowers.

- 5 Promoters that direct expression of nucleic acids in ovules, flowers or seeds are particularly useful in the present invention. As used herein a seed-specific or preferential promoter is one which directs expression specifically or preferentially in seed tissues, such promoters may be, for example, ovule-specific, embryo-specific, endosperm-specific, integument-specific, seed coat-specific, or some combination thereof. Examples
- 10 include a promoter from the ovule-specific BEL1 gene described in Reiser *et al.*, *Cell* 83:735, 1995; (GenBank Accession No. U39944). Other suitable seed specific promoters are derived from the following genes: MAC1 from maize (Sheridan *et al.*, *Genetics* 142:1009, 1996), Cat3 from maize (GenBank Accession No. L05934, Abler *et al.*, *Plant Mol. Biol.* 22:10131, 1993), the gene encoding oleosin 18 kD from maize (GenBank
- 15 Accession No. J05212, Lee *et al.*, *Plant Mol. Biol.* 26:1981, 1994), viviparous-1 from *Arabidopsis* (Genbank Accession No. U93215), the gene encoding oleosin from *Arabidopsis* (Genbank Accession No. Z17657), Atmycl from *Arabidopsis* (Urao *et al.*, *Plant Mol. Biol.* 32:571, 1996), the 2s seed storage protein gene family from *Arabidopsis* (Conceicao *et al.*, *Plant* 5:493, 1994) the gene encoding oleosin 20 kD from *Brassica*
- 20 *napus* (GenBank Accession No. M63985), napA from *Brassica napus* (GenBank Accession No. J02798, Josefsson *et al.*, *J. Biol. Chem.* 262:12196, 1987), the napin gene family from *Brassica napus* (Sjodahl *et al.*, *Planta* 197:264, 1995), the gene encoding the 2S storage protein from *Brassica napus* (Dasgupta *et al.*, *Gene* 133:301, 1993), the genes encoding oleosin A (Genbank Accession No. U09118) and oleosin B (Genbank
- 25 Accession No. U09119) from soybean and the gene encoding low molecular weight sulphur rich protein from soybean (Choi *et al.*, *Mol Gen, Genet.* 246:266, 1995).

Alternatively, particular sequences which provide the promoter with desirable expression characteristics, or the promoter with expression enhancement activity, could be identified and these or similar sequences introduced into the sequences via mutation. It

30 is further contemplated that these sequences can be mutagenized in order to enhance their expression of transgenes in a particular species.

Furthermore, it is contemplated that promoters combining elements from more than one promoter may be useful. For example, U.S. Patent No. 5,491,288 discloses combining a Cauliflower Mosaic Virus promoter with a histone promoter. Thus, the elements from the promoters disclosed herein may be combined with elements from other promoters.

A variety of 5' and 3' transcriptional regulatory sequences are available for use in the present invention. Transcriptional terminators are responsible for the termination of transcription and correct mRNA polyadenylation. The 3' nontranslated regulatory DNA sequence includes from about 50 to about 1,000, or about 100 to about 1,000, nucleotide base pairs and contains plant transcriptional and translational termination sequences. Appropriate transcriptional terminators and those which are known to function in plants include the CaMV 35S terminator, the *tml* terminator, the nopaline synthase terminator, the pea *rbcS* E9 terminator, the terminator for the T7 transcript from the octopine synthase gene of *Agrobacterium tumefaciens*, and the 3' end of the protease inhibitor I or II genes from potato or tomato, although other 3' elements known to those of skill in the art can also be employed. Alternatively, a gamma coixin, oleosin 3 or other terminator from the genus *Coix* can be used.

Non-limiting 3' elements include those from the nopaline synthase gene of *Agrobacterium tumefaciens* (Bevan *et al.*, *Nature* 304:184, 1983), the terminator for the T7 transcript from the octopine synthase gene of *Agrobacterium tumefaciens*, and the 3' end of the protease inhibitor I or II genes from potato or tomato.

As the DNA sequence between the transcription initiation site and the start of the coding sequence, *i.e.*, the untranslated leader sequence, can influence gene expression, one may also wish to employ a particular leader sequence. Non-limiting leader sequences are contemplated to include those which include sequences predicted to direct optimum expression of the attached gene, *i.e.*, to include a consensus leader sequence which may increase or maintain mRNA stability and prevent inappropriate initiation of translation. The choice of such sequences will be known to those of skill in the art in light of the present disclosure. Sequences that are derived from genes that are highly expressed in plants are useful in the present invention.

Other sequences that have been found to enhance gene expression in transgenic plants include intron sequences (*e.g.*, from *Adh1*, *bronzel*, *actin1*, *actin 2* (PCT Publication No. WO 00/760067), or the sucrose synthase intron) and viral leader sequences (*e.g.*, from TMV, MCMV, or AMV). For example, a number of non-translated leader sequences derived from viruses are known to enhance expression. Specifically, leader sequences from Tobacco Mosaic Virus (TMV), Maize Chlorotic Mottle Virus (MCMV), and Alfalfa Mosaic Virus (AMV) have been shown to be effective in enhancing expression (*e.g.*, Gallie *et al.*, *Nuc. Acids Res.* 15:3257, 1987; Skuzeski *et al.*, *Plant Mol. Biol.*, 15: 65, 1990). Other leaders known in the art include but are not limited to: Picornavirus leaders, for example, EMCV leader (Encephalomyocarditis 5 noncoding region) (Elroy-Stein *et al.*, *Proc. Natl. Acad. Sci. USA.* 86:6126, 1989); Potyvirus leaders, for example, TEV leader (Tobacco Etch Virus); MDMV leader (Maize Dwarf Mosaic Virus); Human immunoglobulin heavy-chain binding protein (BiP) leader, (Macejak *et al.*, *Nature* 353:90, 1991); Untranslated leader from the coat protein mRNA of alfalfa mosaic virus (AMV RNA 4), (Jobling *et al.*, *Nature* 325:622, 1987; Tobacco mosaic virus leader (TMV), (Gallie *et al.*, *Plant Cell* 1:301, 1989; and Maize Chlorotic Mottle Virus leader (MCMV) (Lommel *et al.*, *Virology* 181:382, 1991. See also, Della-Cioppa *et al.*, *Plant Physiol.* 84:965, 1987. Regulatory elements such as *Adh* intron 1 (Callis *et al.*, *Genes Dev.* 1:1183, 1987), sucrose synthase intron (Vasil *et al.*, *Mol. Microbiol.* 3:371, 1989) or TMV omega element (Gallie *et al.*, *Plant Cell* 1:301, 1989), may further be included where desired. Non-limiting examples of enhancers include elements from the CaMV 35S promoter, octopine synthase genes (Ellis *et al.*, *EMBO J.*, 6:3203, 1987), the rice actin I gene, the maize alcohol dehydrogenase gene (Callis *et al.*, *Genes Dev.* 1:1183, 1987), the maize shrunken I gene (Vasil *et al.*, *Mol. Microbiol.* 3:371, 1989), TMV Omega element (Gallie *et al.*, *Plant Cell* 1:301, 1989) and promoters from non-plant eukaryotes (*e.g.* yeast; Ma *et al.*, *Nature* 334:631, 1988).

A host cell is any type of cell including, without limitation, a bacterial cell, a yeast cell, a plant cell, an insect cell, and a mammalian cell. Numerous such cells are commercially available, for example, from the American Type Culture Collection, Manassas, Virginia.

In certain embodiments, the cell is a plant cell, which can be regenerated to form

a transgenic plant. Thus, the present invention provides a transformed (transgenic) plant cell, *in planta* or *ex planta*, including a transformed plastid or other organelle, *e.g.*, nucleus, mitochondria or chloroplast. As used herein, by "transgenic plant" is a plant having one or more plant cells that contain an exogenous nucleic acid molecule (*e.g.*, a
 5 nucleic acid molecule encoding a cell proliferation-related polypeptide of the invention).

The present invention may be used for transformation of any plant species, including, but not limited to, cells from corn (*Zea mays*), *Brassica* sp. (*e.g.*, *B. napus*, *B. rapa*, *B. juncea*), particularly those *Brassica* species useful as sources of seed oil, alfalfa (*Medicago sativa*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor*,
 10 *Sorghum vulgare*), millet (*e.g.*, pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*)), sunflower (*Helianthus annuus*), safflower (*Carthamus tinctorius*), wheat (*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanuts (*Arachis hypogaea*), cotton (*Gossypium barbadense*, *Gossypium hirsutum*),
 15 sweet potato (*Ipomoea batatas*), cassava (*Manihot esculenta*), coffee (*Cofea* spp.), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (*Persea utilane*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*),
 20 macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), sugarcane (*Saccharum* spp.), oats, duckweed (*Lemna*), barley, vegetables, ornamentals, and conifers.

Duckweed (*Lemna*, see PCT Publication No. WO 00/07210) includes members of the family *Lemnaceae*. There are known four genera and 34 species of duckweed as
 25 follows: genus *Lemna* (*L. aequinoctialis*, *L. disperma*, *L. ecuadoriensis*, *L. gibba*, *L. japonica*, *L. minor*, *L. miniscula*, *L. obscura*, *L. perpusilla*, *L. tenera*, *L. trisulca*, *L. turionifera*, *L. valdiviana*); genus *Spirodela* (*S. intermedia*, *S. polyrrhiza*, *S. punctata*); genus *Woffia* (*Wa. Angusta*, *Wa. Arrhiza*, *Wa. Australina*, *Wa. Borealis*, *Wa. Brasiliensis*, *Wa. Columbiana*, *Wa. Elongata*, *Wa. Globosa*, *Wa. Microscopica*, *Wa. Neglecta*) and
 30 genus *Wofilla* (*Wl. utila*, *Wl. utilanen*, *Wl. gladiata*, *Wl. utila*, *Wl. lingulata*, *Wl. repunda*, *Wl. rotunda*, and *Wl. neotropica*). Any other genera or species of *Lemnaceae*,

if they exist, are also aspects of the present invention. *Lemna gibba*, *Lemna minor*, and *Lemna miniscula* are included in the invention, as are *Lemna minor* and *Lemna miniscula*. *Lemna* species can be classified using the taxonomic scheme described by Landolt, Biosystematic Investigation on the Family of Duckweeds: The family of Lemnaceae – A Monograph Study, Geobotanischen Institut ETH, Stiftung Rubel, Zurich (1986)).

Vegetables within the scope of the invention include tomatoes (*Lycopersicon esculentum*), lettuce (e.g., *Lactuca sativa*), green beans (*Phaseolus vulgaris*), lima beans (*Phaseolus limensis*), peas (*Lathyrus* spp.), and members of the genus *Cucumis* such as cucumber (*C. sativus*), cantaloupe (*C. cantalupensis*), and musk melon (*C. melo*). Ornamentals include azalea (*Rhododendron* spp.), hydrangea (*Macrophylla hydrangea*), hibiscus (*Hibiscus rosasanensis*), roses (*Rosa* spp.), tulips (*Tulipa* spp.), daffodils (*Narcissus* spp.), petunias (*Petunia hybrida*), carnation (*Dianthus caryophyllus*), poinsettia (*Euphorbia pulcherrima*), and chrysanthemum. Conifers that may be employed in practicing the present invention include, for example, pines such as loblolly pine (*Pinus taeda*), slash pine (*Pinus elliotii*), ponderosa pine (*Pinus ponderosa*), lodgepole pine (*Pinus contorta*), and Monterey pine (*Pinus radiata*), Douglas-fir (*Pseudotsuga menziesii*), Western hemlock (*Tsuga utilane*), Sitka spruce (*Picea glauca*), redwood (*Sequoia sempervirens*), true firs such as silver fir (*Abies amabilis*) and balsam fir (*Abies balsamea*); and cedars such as Western red cedar (*Thuja plicata*) and Alaska yellow-cedar (*Chamaecyparis nootkatensis*). Leguminous plants include beans and peas. Beans include guar, locust bean, fenugreek, soybean, garden beans, cowpea, mungbean, lima bean, fava bean, lentils, chickpea, etc. Legumes include, but are not limited to, *Arachis*, e.g., peanuts, *Vicia*, e.g., crown vetch, hairy vetch, adzuki bean, mung bean, and chickpea, *Lupinus*, e.g., lupine, trifolium, *Phaseolus*, e.g., common bean and lima bean, *Pisum*, e.g., field bean, *Melilotus*, e.g., clover, *Medicago*, e.g., alfalfa, Lotus, e.g., trefoil, lens, e.g., lentil, and false indigo. Non-limiting forage and turf grass for use in the methods of the invention include alfalfa, orchard grass, tall fescue, perennial ryegrass, creeping bent grass, and redtop.

Other plants within the scope of the invention include *Acacia*, aneth, artichoke, arugula, blackberry, canola, cilantro, clementines, escarole, eucalyptus, fennel, grapefruit,

honey dew, jicama, kiwifruit, lemon, lime, mushroom, nut, okra, orange, parsley, persimmon, plantain, pomegranate, poplar, radiata pine, radicchio, Southern pine, sweetgum, tangerine, triticale, vine, yams, apple, pear, quince, cherry, apricot, melon, hemp, buckwheat, grape, raspberry, chenopodium, blueberry, nectarine, peach, plum, strawberry, watermelon, eggplant, pepper, cauliflower, Brassica, *e.g.*, broccoli, cabbage, 5 utilian sprouts, onion, carrot, leek, beet, broad bean, celery, radish, pumpkin, endive, gourd, garlic, snapbean, spinach, squash, turnip, utilane, and zucchini.

Ornamental plants within the scope of the invention include impatiens, Begonia, Pelargonium, Viola, Cyclamen, Verbena, Vinca, Tagetes, Primula, Saint Paulia, 10 Agertum, Amaranthus, Antihirrhinum, Aquilegia, Cineraria, Clover, Cosmo, Cowpea, Dahlia, Datura, Delphinium, Gerbera, Gladiolus, Gloxinia, Hippeastrum, Mesembryanthemum, Salpiglossos, and Zinnia.

In certain embodiments, transgenic plants of the present invention are crop plants and in particular cereals (for example, corn, alfalfa, sunflower, rice, Brassica, canola, 15 soybean, barley, soybean, sugarbeet, cotton, safflower, peanut, sorghum, wheat, millet, tobacco), or corn, rice and soybean.

The present invention also provides a transgenic plants, a seed from such a plant and progeny plants from such a plant including hybrids and inbreds. In some 20 embodiments, transgenic plants are transgenic maize, soybean, barley, alfalfa, sunflower, canola, soybean, cotton, peanut, sorghum, tobacco, sugarbeet, rice, wheat, rye, turfgrass, millet, sugarcane, tomato, or potato.

A transformed (transgenic) plant of the invention includes plants, the genome of which is augmented by a nucleic acid molecule of the invention, or in which the 25 corresponding gene has been disrupted, *e.g.*, to result in a loss, a decrease or an alteration, in the function of the product encoded by the gene, which plant may also have increased yields and/or produce a better-quality product than the corresponding wild-type plant. The nucleic acid molecules of the invention are thus useful for targeted gene disruption, as well as markers and probes.

The invention also provides a method of plant breeding, *e.g.*, to prepare a crossed 30 fertile transgenic plant. The method comprises crossing a fertile transgenic plant comprising a particular nucleic acid molecule of the invention with itself or with a second

plant, e.g., one lacking the particular nucleic acid molecule, to prepare the seed of a crossed fertile transgenic plant comprising the particular nucleic acid molecule. The seed is then planted to obtain a crossed fertile transgenic plant. The plant may be a monocot or a dicot. In a particular embodiment, the plant is a cereal plant.

5 The crossed fertile transgenic plant may have the particular nucleic acid molecule inherited through a female parent or through a male parent. The second plant may be an inbred plant. The crossed fertile transgenic may be a hybrid. Also included within the present invention are seeds of any of these crossed fertile transgenic plants.

10 Transformation of plants can be undertaken with a single DNA molecule or multiple DNA molecules (i.e., co-transformation), and both these techniques are suitable for use with the expression cassettes of the present invention. Numerous transformation vectors are available for plant transformation, and the expression cassettes of this invention can be used in conjunction with any such vectors. The selection of vector will depend upon the transformation technique and the target species for transformation.

15 A variety of techniques are available and known for introduction of nucleic acid molecules and expression cassettes comprising such nucleic acid molecules into a plant cell host. These techniques generally include transformation with DNA employing *A. tumefaciens* or *A. rhizogenes* as the transforming agent, liposomes, PEG precipitation, electroporation, DNA injection, direct DNA uptake, microprojectile bombardment, particle acceleration, and the like (See, for example, EP 295959 and EP 138341) (see below). However, cells other than plant cells may be transformed with the expression cassettes of the invention. The general descriptions of plant expression vectors and reporter genes, and *Agrobacterium* and *Agrobacterium*-mediated gene transfer, can be found in Gruber *et al.*, *Vectors for Plant Transformation*, in Methods in Plant Molecular
20 Biology, Glich *et al.*, eds, pp. 89-119, CRC Press, 1993.

25 Expression vectors containing genomic or synthetic fragments can be introduced into protoplasts or into intact tissues or isolated cells. In some embodiments, expression vectors are introduced into intact tissue. "Plant tissue" includes differentiated and undifferentiated tissues or plants, including but not limited to roots, stems, shoots, leaves,
30 pollen, seeds, tumor tissue and various forms of cells and culture such as single cells, protoplast, embryos, and callus tissue. The plant tissue may be in plants or in organ,

tissue or cell culture. General methods of culturing plant tissues are provided for example by Maki *et al.*, Methods in Plant Molecular Biology, Glich *et al.*, eds, pp. 67-88, CRC Press, 1993; and by Phillips *et al.* in *Corn and Corn Improvement*, 3rd ed, Sprague *et al.*, eds., Amer. Soc of Agronomy, 1988. In some embodiments, expression vectors are introduced into maize or other plant tissues using a direct gene transfer method such as microprojectile-mediated delivery, DNA injection, electroporation and the like. In some
5 embodiments, expression vectors are introduced into plant tissues using the microprojectile media delivery with the biolistic device (see, for example, Tomes *et al.*, *Plant Cell, Tissue and Organ Culture: Fundamental Methods*, Springer-Verlag, 1995).
10 The vectors of the invention can not only be used for expression of structural genes but may also be used in exon-trap cloning, or promoter trap procedures to detect differential gene expression in varieties of tissues (Lindsey *et al.*, *Transgen. Res.* 2: 3347, 1993; Auch and Reth *et al.*, *Nuc. Acids Res.* 18:6743, 1990).

In some embodiments, the binary type vectors of Ti and Ri plasmids of
15 *Agrobacterium spp.* Ti-derived vectors may be used transform a wide variety of higher plants, including monocotyledonous and dicotyledonous plants, such as soybean, cotton, rape, tobacco, and rice (Pacciotti *et al.*, *Bio/Technology* 3:241, 1985; Byrne *et al.*, *Plant Cell Tissue Org Culture* 8:3, 1987; Sukhapinda *et al.*, *Plant Mol. Biol.* 8:209, 1987; Lorz *et al.*, *Mol. Gen. Genet.* 199:178, 1985; Potrykus, *Trends Biotech.* 7:269 1985; Park *et al.*,
20 *J. Plant Biol.*, 38:365, 1985; Hiei *et al.*, *Plant J.* 6:271, 1994). The use of T-DNA to transform plant cells has received extensive study and is amply described (European Patent Application No. EP 120516; Hoekema, in *The Binary Plant Vector System*, Offset-drukkerij Kanters B.V., 1985; Knauf, *et al.*, *Analysis of Host Range Expression by Agrobacterium*, in Molecular Genetics of the Bacteria-Plant Interaction, Puhler, ed.,
25 Springer-Verlag, 1983; and An *et al.*, *EMBO J.* 4:277, 1985). For introduction into plants, the chimeric genes of the invention can be inserted into binary vectors as described in the examples.

Other transformation methods are available to those skilled in the art, such as direct uptake of foreign DNA constructs (see European Patent Application No. EP
30 295959), techniques of electroporation (Fromm *et al.*, *Nature* 319:791 1986) or high velocity ballistic bombardment with metal particles coated with the nucleic acid

constructs (Kline *et al.*, *Nature* 327:70, 1987, and U.S. Patent No. 4,945,050). Once transformed, the cells can be regenerated by those skilled in the art. Of particular relevance are the recently described methods to transform foreign genes into commercially important crops, such as rapeseed (De Block *et al.*, *Plant Physiol.* 91:694, 1989), sunflower (Everett *et al.*, *Bio/Technology* 5:1201, 1987), soybean (McCabe *et al.*, *Bio/Technology* 6:923, 1988; Hinchee *et al.*, *Bio/Technology* 6:915, 1988; Chee *et al.*, *Plant Physiol.* 91:1212, 1989; Christou *et al.*, *Proc. Natl. Acad. Sci. USA* 86:7500, 1989; European Patent Application No. EP 301749), rice (Hiei *et al.*, *Plant J.* 6:271, 1994), and corn (Gordon Kamm *et al.*, *Plant Cell* 2:603, 1990; Fromm *et al.*, *Bio/Technology* 8:833, 1990).

Of course, the choice of method might depend on the type of plant, *i.e.*, monocotyledonous or dicotyledonous, targeted for transformation. Suitable methods of transforming plant cells include, but are not limited to, microinjection (Crossway *et al.*, *Bio/Techniques* 4:320, 1986), electroporation (Riggs *et al.*, *Proc. Natl. Acad. Sci. USA* 83:5602, 1986), *Agrobacterium*-mediated transformation (Hinchee *et al.*, *Bio/Technology* 6:915, 1988), direct gene transfer (Paszowski *et al.*, *EMBO J.* 3:2717, 1984), and ballistic particle acceleration using devices available from Agracetus, Inc., Madison, Wis. And BioRad, Hercules, Calif. (see, for example, Sanford *et al.*, U.S. Pat. No. 4,945,050; and McCabe *et al.*, *Bio/Technology* 6:923, 1988). Also see, Weissinger *et al.*, *Ann. Rev. Genet.* 22:421, 1988; Sanford *et al.*, *Particulate Sci. Tech.* 5:27, 1987 (onion); Christou *et al.*, *Plant Physiol.* 87:671, 1988 (soybean); McCabe *et al.*, *Bio/Technology* 6:923, 1988 (soybean); Datta *et al.*, *Bio/Technology* 8: 736, 1990 (rice); Klein *et al.*, *Bio/Technology* 6:559, 1988 (maize); Fromm *et al.*, *Bio/Technology* 8: 833, 1990 (maize); and Gordon-Kamm *et al.*, *Plant Cell* 2:603, 1990 (maize); Svab *et al.*, *Proc. Natl. Acad. Sci. USA* 87:8526, 1990 (tobacco chloroplast); Koziel *et al.*, *Biotechnology* 11:194, 1993 (maize); Shimamoto *et al.*, *Nature* 338:274, 1989 (rice); Christou *et al.*, *Biotechnology* 9:957, 1991 (rice); European Patent Application EP 0 332 581 (orchardgrass and other Pooideae); Vasil *et al.*, *Biotechnology* 11:1553, 1993 (wheat); Weeks *et al.*, *Plant Physiol.* 102:1077, 1993 (wheat). In one embodiment, the protoplast transformation method for maize is employed (European Patent Application EP 0 292 435, U. S. Pat. No. 5,350,689).

In another embodiment, a nucleotide sequence of the present invention is directly transformed into the plastid genome. Plastid transformation technology is extensively described in U.S. Patent Nos. 5,451,513, 5,545,817, and 5,545,818, in PCT Publication No. WO 95/16783, and in McBride *et al.*, *Proc. Natl. Acad. Sci. USA* 91:7301, 1994. The basic technique for chloroplast transformation involves introducing regions of cloned plastid DNA flanking a selectable marker together with the gene of interest into a suitable target tissue, *e.g.*, using biolistics or protoplast transformation (*e.g.*, calcium chloride or PEG mediated transformation). The 1 to 1.5 kb flanking regions, termed targeting sequences, facilitate orthologous recombination with the plastid genome and thus allow the replacement or modification of specific regions of the plastome. Initially, point mutations in the chloroplast 16S rRNA and rps12 genes conferring resistance to spectinomycin and/or streptomycin are utilized as selectable markers for transformation (Svab *et al.*, *Proc. Natl. Acad. Sci. USA* 87:8526, 1990; Staub *et al.*, *Plant Cell* 4:39, 1992). This resulted in stable homoplasmic transformants at a frequency of approximately one per 100 bombardments of target leaves. The presence of cloning sites between these markers allowed creation of a plastid targeting vector for introduction of foreign genes (Staub *et al.*, *EMBO J.* 12:601, 1993). Substantial increases in transformation frequency are obtained by replacement of the recessive rRNA or r-protein antibiotic resistance genes with a dominant selectable marker, the bacterial *aadA* gene encoding the spectinomycin-detoxifying enzyme aminoglycoside-3N-adenyltransferase (Svab *et al.*, *EMBO J.* 12:601, 1993). Other selectable markers useful for plastid transformation are known in the art and encompassed within the scope of the invention. Typically, approximately 15-20 cell division cycles following transformation are required to reach a homoplastidic state. Plastid expression, in which genes are inserted by orthologous recombination into all of the several thousand copies of the circular plastid genome present in each plant cell, takes advantage of the enormous copy number advantage over nuclear-expressed genes to permit expression levels that can readily exceed 10% of the total soluble plant protein. In one embodiment, a nucleotide sequence of the present invention is inserted into a plastid targeting vector and transformed into the plastid genome of a desired plant host. Plants homoplasmic for plastid genomes

containing a nucleotide sequence of the present invention are obtained, and are preferentially capable of high expression of the nucleotide sequence.

5 *Agrobacterium tumefaciens* cells containing a vector comprising an expression cassette of the present invention, wherein the vector comprises a Ti plasmid, are useful in methods of making transformed plants. Plant cells are infected with an *Agrobacterium tumefaciens* as described above to produce a transformed plant cell, and then a plant is regenerated from the transformed plant cell. Numerous *Agrobacterium* vector systems useful in carrying out the present invention are known.

For example, vectors are available for transformation using *Agrobacterium*
10 *tumefaciens*. These typically carry at least one T-DNA border sequence and include vectors such as pBIN19 (Bevan, *Nuc. Acids Res.* 12:8711, 1984). In one non-limiting embodiment, the expression cassettes of the present invention may be inserted into either of the binary vectors pCIB200 and pCIB2001 for use with *Agrobacterium*. These vector cassettes for *Agrobacterium*-mediated transformation were constructed in the following
15 manner. PTJS75kan was created by NarI digestion of pTJS75 (Schmidhauser & Helinski, *J. Bacteriol.* 164:446, 1985) allowing excision of the tetracycline-resistance gene, followed by insertion of an AccI fragment from pUC4K carrying an NPTII (Messing & Vierra, *Gene* 19:259, 1982; Bevan *et al.*, *Nature* 304:184, 1983; McBride *et al.*, *Plant Mol. Biol.* 14:266, 1990). XhoI linkers are ligated to the EcoRV fragment of pCIB7
20 which contains the left and right T-DNA borders, a plant selectable nos/nptII chimeric gene and the pUC polylinker (Rothstein *et al.*, *Gene* 53:153, 1987), and the XhoI-digested fragment was cloned into SalI-digested pTJS75kan to create pCIB200 (see also European Patent Application No. EP 0 332 104, example 19). PCIB200 contains the following unique polylinker restriction sites: EcoRI, SstI, KpnI, BglII, XbaI, and SalI.
25 The plasmid pCIB2001 is a derivative of pCIB200 which was created by the insertion into the polylinker of additional restriction sites. Unique restriction sites in the polylinker of pCIB2001 are EcoRI, SstI, KpnI, BglII, XbaI, SalI, MluI, BclI, AvrII, ApaI, HpaI, and StuI. PCIB2001, in addition to containing these unique restriction sites also has plant and bacterial kanamycin selection, left and right T-DNA borders for *Agrobacterium*-mediated
30 transformation, the RK2-derived trfA function for mobilization between *E. coli* and other hosts, and the OriT and OriV functions also from RK2. The pCIB2001 polylinker is

suitable for the cloning of plant expression cassettes containing their own regulatory signals.

An additional vector useful for *Agrobacterium*-mediated transformation is the binary vector pCIB 10, which contains a gene encoding kanamycin resistance for selection in plants, T-DNA right and left border sequences and incorporates sequences from the wide host- range plasmid pRK252 allowing it to replicate in both *E. coli* and *Agrobacterium*. Its construction is described by Rothstein *et al.*, *Gene* 53:153, 1987. Various derivatives of pCIB10 have been constructed which incorporate the gene for hygromycin B phosphotransferase described by Gritz *et al.*, *Gene*, 25:179, 1983. These derivatives enable selection of transgenic plant cells on hygromycin only (pCIB743), or hygromycin and kanamycin (pCIB715, pCIB717).

Methods using either a form of direct gene transfer or *Agrobacterium*-mediated transfer usually, but not necessarily, are undertaken with a selectable marker which may provide resistance to an antibiotic (*e.g.*, kanamycin, hygromycin or methotrexate) or a herbicide (*e.g.*, phosphinothricin). The choice of selectable marker for plant transformation is not, however, critical to the invention.

For certain plant species, different antibiotic or herbicide selection markers may be employed. Selection markers used routinely in transformation include the nptII gene which confers resistance to kanamycin and related antibiotics (Messing & Vierra, *Gene* 19:259, 1982; Bevan *et al.*, *Nature* 304:184, 1983), the bar gene which confers resistance to the herbicide phosphinothricin (White *et al.*, *Nuc. Acids Res.* 18:1062, 1990, Spencer *et al.*, *Theor. Appl. Genet.* 79:625, 1990), the hph gene which confers resistance to the antibiotic hygromycin (Blochinger and Diggelmann *Mol. Cell. Biol.* 4:2929, 1984), and the dhfr gene, which confers resistance to methotrexate (Bourouis *et al.*, *EMBO J.* 2:1099 1983).

Selection markers resulting in positive selection, such as a phosphomannose isomerase gene, as described in PCT Publication No. WO 93/05163, are also used. Other genes to be used for positive selection are described in PCT Publication No. WO 94/20627 and encode xyloisomerases and phosphomanno-isomerases such as mannose-6-phosphate isomerase and mannose-1-phosphate isomerase; phosphomanno mutase; mannose epimerases such as those which convert carbohydrates to mannose or mannose to

carbohydrates such as glucose or galactose; phosphatases such as mannose or xylose phosphatase, mannose-6-phosphatase and mannose-1-phosphatase, and permeases which are involved in the transport of mannose, or a derivative, or a precursor thereof into the cell. The agent which reduces the toxicity of the compound to the cells is typically a glucose derivative such as methyl-3-O-glucose or phloridzin. Transformed cells are identified without damaging or killing the non-transformed cells in the population and without co-introduction of antibiotic or herbicide resistance genes. As described in PCT Publication No. WO 93/05163, in addition to the fact that the need for antibiotic or herbicide resistance genes is eliminated, it has been shown that the positive selection method is often far more efficient than traditional negative selection.

One vector useful for direct gene transfer techniques in combination with selection by the herbicide Basta (or phosphinothricin) is pCIB3064. This vector is based on the plasmid pCIB246, which comprises the CaMV 35S promoter in operational fusion to the *E. coli* GUS gene and the CaMV 35S transcriptional terminator and is described in PCT Publication No. WO 93/07278. One gene useful for conferring resistance to phosphinothricin is the bar gene from *Streptomyces viridochromogenes* (Thompson *et al.*, *EMBO J.* 6:2519, 1987). This vector is suitable for the cloning of plant expression cassettes containing their own regulatory signals.

An additional transformation vector is pSOG35 which utilizes the *E. coli* gene dihydrofolate reductase (DHFR) as a selectable marker conferring resistance to methotrexate. PCR was used to amplify the 35S promoter (about 800 bp), intron 6 from the maize Adh1 gene (about 550 bp) and 18 bp of the GUS untranslated leader sequence from pSOG10. A 250 bp fragment encoding the *E. coli* dihydrofolate reductase type II gene was also amplified by PCR and these two PCR fragments are assembled with a SacI-PstI fragment from pBI221 (Clontech) which comprised the pUC19 vector backbone and the nopaline synthase terminator. Assembly of these fragments generated pSOG19 which contains the 35S promoter in fusion with the intron 6 sequence, the GUS leader, the DHFR gene and the nopaline synthase terminator. Replacement of the GUS leader in pSOG19 with the leader sequence from Maize Chlorotic Mottle Virus check (MCMV) generated the vector pSOG35. pSOG19 and pSOG35 carry the pUC-derived

gene for ampicillin resistance and have HindIII, SphI, PstI and EcoRI sites available for the cloning of foreign sequences.

5 Binary backbone vector pNOV2117 contains the T-DNA portion flanked by the right and left border sequences, and including the Positech™ (Syngenta) plant selectable marker and the "candidate gene" gene expression cassette. The Positech™ plant selectable marker confers resistance to mannose and in this instance consists of the maize ubiquitin promoter driving expression of the PMI (phosphomannose isomerase) gene, followed by the cauliflower mosaic virus transcriptional terminator.

10 Transgenic plant cells are then placed in an appropriate selective medium for selection of transgenic cells which are then grown to callus. Shoots are grown from callus and plantlets generated from the shoot by growing in rooting medium. The various constructs normally are joined to a marker for selection in plant cells. Conveniently, the marker may be resistance to a biocide (particularly an antibiotic, such as kanamycin, G418, bleomycin, hygromycin, chloramphenicol, herbicide, or the like). The particular
15 marker used allows for selection of transformed cells as compared to cells lacking the DNA which has been introduced. Components of DNA constructs including transcription cassettes of this invention are prepared from sequences which are native (endogenous) or foreign (exogenous) to the host. By "foreign," is meant that the sequence is not found in the wild-type host into which the construct is introduced. Heterologous constructs will
20 contain at least one region which is not native to the gene from which the transcription-initiation-region is derived.

To confirm the presence of the transgenes in transgenic cells and plants, a variety of assays may be performed. Such assays include, for example, "molecular biological" assays well known to those of skill in the art, such as Southern and Northern blotting, *in situ* hybridization and nucleic acid-based amplification methods such as PCR or RT-PCR; "biochemical" assays, such as detecting the presence of a protein product, *e.g.*, by immunological means (ELISAs and Western blots) or by enzymatic function; plant part assays, such as seed assays; and also, by analyzing the phenotype of the whole regenerated plant, *e.g.*, for disease or pest resistance.

30 DNA may be isolated from cell lines or any plant parts to determine the presence of the preselected nucleic acid segment through the use of techniques well known to

those skilled in the art. Note that intact sequences will not always be present, presumably due to rearrangement or deletion of sequences in the cell.

The presence of nucleic acid elements introduced through the methods of this invention may be determined by polymerase chain reaction (PCR). Using this technique
5 discrete fragments of nucleic acid are amplified and detected by gel electrophoresis. This type of analysis permits one to determine whether a preselected nucleic acid segment is present in a stable transformant. It is contemplated that using PCR techniques it would be possible to clone fragments of the host genomic DNA adjacent to an introduced preselected DNA segment.

10 Positive proof of DNA integration into the host genome and the independent identities of transformants may be determined using the technique of Southern hybridization. Using this technique specific DNA sequences that are introduced into the host genome and flanking host DNA sequences can be identified. Hence the Southern hybridization pattern of a given transformant serves as an identifying characteristic of
15 that transformant. In addition it is possible through Southern hybridization to demonstrate the presence of introduced preselected DNA segments in high molecular weight DNA, *i.e.*, confirm that the introduced preselected DNA segment has been integrated into the host cell genome. The technique of Southern hybridization provides information that is obtained using PCR, *e.g.*, the presence of a preselected DNA segment,
20 but also demonstrates integration into the genome and characterizes each individual transformant.

It is contemplated that using the techniques of dot or slot blot hybridization which are modifications of Southern hybridization techniques, the same information that is derived from PCR could be obtained, *e.g.*, the presence of a preselected DNA segment.

25 Both PCR and Southern hybridization techniques can be used to demonstrate transmission of a preselected DNA segment to progeny. In most instances the characteristic Southern hybridization pattern for a given transformant will segregate in progeny as one or more Mendelian genes (Spencer *et al.*, *Theor. Appl. Genet.* 79:625, 1992); Laursen *et al.*, *Plant Mol. Biol.* 24:51, 1994) indicating stable inheritance of the
30 gene. The nonchimeric nature of the callus and the parental transformants (R_0) was suggested by germline transmission and the identical Southern blot hybridization patterns

and intensities of the transforming DNA in callus, R₀ plants and R₁ progeny that segregated for the transformed gene.

Whereas DNA analysis techniques may be conducted using DNA isolated from any part of a plant, RNA may only be expressed in particular cells or tissue types and hence it will be necessary to prepare RNA for analysis from these tissues. PCR techniques may also be used for detection and quantitation of RNA produced from introduced preselected DNA segments. In this application of PCR it is first necessary to reverse transcribe RNA into DNA, using enzymes such as reverse transcriptase, and then through the use of conventional PCR techniques amplify the DNA. In most instances PCR techniques, while useful, will not demonstrate integrity of the RNA product. Further information about the nature of the RNA product may be obtained by Northern blotting. This technique demonstrates the presence of an RNA species and give information about the integrity of that RNA. The presence or absence of an RNA species can also be determined using dot or slot blot Northern hybridizations. These techniques are modifications of Northern blotting and will only demonstrate the presence or absence of an RNA species.

Thus, Southern blotting and PCR may be used to detect the preselected DNA segment in question. Expression may be evaluated by specifically identifying the protein products of the introduced preselected DNA segments or evaluating the phenotypic changes brought about by their expression.

Assays for the production and identification of specific proteins may make use of physical-chemical, structural, functional, or other properties of the proteins. Unique physical-chemical or structural properties allow the proteins to be separated and identified by electrophoretic procedures, such as native or denaturing gel electrophoresis or isoelectric focusing, or by chromatographic techniques such as ion exchange or gel exclusion chromatography. The unique structures of individual proteins offer opportunities for use of specific antibodies to detect their presence in formats such as an ELISA assay. Combinations of approaches may be employed with even greater specificity such as Western blotting in which antibodies are used to locate individual gene products that have been separated by electrophoretic techniques. Additional techniques may be employed to absolutely confirm the identity of the product of interest

such as evaluation by amino acid sequencing following purification. Although these are among the most commonly employed, other procedures may be additionally used.

Assay procedures may also be used to identify the expression of proteins by their functionality, especially the ability of enzymes to catalyze specific chemical reactions
5 involving specific substrates and products. These reactions may be followed by providing and quantifying the loss of substrates or the generation of products of the reactions by physical or chemical procedures. Examples are as varied as the enzyme to be analyzed.

Very frequently the expression of a gene product is determined by evaluating the
10 phenotypic results of its expression. These assays also may take many forms including but not limited to analyzing changes in the chemical composition, morphology, or physiological properties of the plant. Morphological changes may include greater stature or thicker stalks. Most often changes in response of plants or plant parts to imposed treatments are evaluated under carefully controlled conditions termed bioassays.

15 The compositions of the invention include plant nucleic acid molecules, and the amino acid sequences for the polypeptides or partial-length polypeptides encoded by the nucleic acid molecule which comprises an open reading frame. These sequences can be employed to alter expression of a particular gene corresponding to the open reading frame by decreasing or eliminating expression of that plant gene or by overexpressing a
20 particular gene product. Methods of this embodiment of the invention include stably transforming a plant with the nucleic acid molecule of the invention which includes an open reading frame operably linked to a promoter capable of driving expression of that open reading frame (sense or antisense) in a plant cell. By "portion" or "fragment", as it relates to a nucleic acid molecule which comprises an open reading frame or a fragment
25 thereof encoding a partial-length polypeptide having the activity of the full length polypeptide, is meant a sequence having at least 80 nucleotides, or at least 150 nucleotides, or at least 400 nucleotides. If not employed for expressing, a "portion" or "fragment" means at least 9, or 12, or 15, or at least 20, consecutive nucleotides, *e.g.*, probes and primers (oligonucleotides), corresponding to the nucleotide sequence of the
30 nucleic acid molecules of the invention. Thus, to express a particular gene product, the method comprises introducing to a plant, plant cell, or plant tissue an expression cassette

comprising a promoter linked to an open reading frame so as to yield a transformed differentiated plant, transformed cell or transformed tissue. Transformed cells or tissue can be regenerated to provide a transformed differentiated plant. The transformed differentiated plant or cells thereof expresses the open reading frame in an amount that
5 alters the amount of the gene product in the plant or cells thereof, which product is encoded by the open reading frame. The present invention also provides a transformed plant prepared by the method, progeny and seed thereof.

The invention further includes a nucleotide sequence which is complementary to one (hereinafter "test" sequence) which hybridizes under stringent conditions with a
10 nucleic acid molecule of the invention as well as RNA which is transcribed from the nucleic acid molecule. When the hybridization is performed under stringent conditions, either the test or nucleic acid molecule of invention may be supported, *e.g.*, on a membrane or DNA chip. Thus, either a denatured test or nucleic acid molecule of the invention is first bound to a support and hybridization is effected for a specified period of
15 time at a temperature of, *e.g.*, between 55 and 70°C, in double strength citrate buffered saline (SC) containing 0.1% SDS followed by rinsing of the support at the same temperature but with a buffer having a reduced SC concentration. Depending upon the degree of stringency required such reduced concentration buffers are typically single strength SC containing 0.1% SDS, half strength SC containing 0.1% SDS and one-tenth
20 strength SC containing 0.1% SDS.

In a further embodiment, the present invention provides a transformed plant host cell, or one obtained through breeding, capable of over-expressing, under-expressing, or having a knock out of amino acid genes and/or their gene products. The plant cell is transformed with at least one such expression vector wherein the plant host cell can be
25 used to regenerate plant tissue or an entire plant, or seed there from, in which the effects of expression, including overexpression or underexpression, of the introduced sequence or sequences can be measured *in vitro* or *in planta*.

In another aspect, the invention features an isolated cell proliferation-related polypeptide, wherein the polypeptide binds to a fragment of a protein selected from the
30 group consisting of OsE2F1, Os018989-4003, OsE2F2, OsS49462, OsCYCOS2, OsMADS45, OsRAP1B, OsMADS6, OsFDRMADS8, OsMADS3, OsMADS5,

OsMADS15, OsHOS59, OsGF14-c, OsDAD1, Os006819-2510, OsCRTC, OsSGT1, OsPN31085, OsCHIB1, OsCS, OsPP2A-2, and OsCAA90866. In some embodiments, the invention features an isolated polypeptide comprising or consisting of an amino acid sequence substantially similar to the amino acid sequence of an isolated cell proliferation-related polypeptide of the invention.

Because the proteins of the invention have a roll in cell proliferation, in certain embodiments, a cell introduced with a nucleic acid molecule of the invention has a different cell proliferation rate as compared to a cell not introduced with the nucleic acid molecule.

In another aspect, the invention features a method for modulating the proliferation of a plant cell comprising introducing an isolated nucleic acid molecule encoding a cell proliferation-related polypeptide into the plant cell, wherein the polypeptide binds to a fragment of a protein selected from the group consisting of OsE2F1, Os018989-4003, OsE2F2, OsS49462, OsCYCOS2, OsMADS45, OsRAP1B, OsMADS6, OsFDRMADS8, OsMADS3, OsMADS5, OsMADS15, OsHOS59, OsGF14-c, OsDAD1, Os006819-2510, OsCRTC, OsSGT1, OsERP, OsCHIB1, OsCS, OsPP2A-2, and OsCAA90866, wherein the polypeptide is expressed by the cell.

In another aspect, the invention features a method for modulating the proliferation of a plant cell comprising introducing an isolated nucleic acid molecule encoding a cell proliferation-related polypeptide into the plant cell, wherein the polypeptide binds to a fragment of a protein selected from the group consisting of OsE2F1, Os018989-4003, OsE2F2, OsS49462, OsCYCOS2, OsMADS45, OsRAP1B, OsMADS6, OsFDRMADS8, OsMADS3, OsMADS5, OsMADS15, OsHOS59, OsGF14-c, OsDAD1, Os006819-2510, OsCRTC, OsSGT1, OsERP, OsCHIB1, OsCS, OsPP2A-2, and OsCAA90866, wherein expression of the polypeptide encoded by the nucleic acid molecule is reduced in the cell.

As discussed herein, all of the cell proliferation-related proteins described herein affect cell proliferation, either under normal conditions, under adverse conditions (*e.g.*, when the plant is exposed to stress (biotic or abiotic), or when the plant is developing and differentiating. Accordingly, by changing the amount of a cell proliferation-related protein of the invention in a plant cell, the proliferation of that plant cell can be modulated.

In some situations, increasing expression of a cell proliferation-related protein of the invention in a cell will cause that cell to increase its rate of poliferation, either alone or in response to some stimuli (*e.g.*, stress or growth hormone). In other situations, increasing expression of a cell proliferation-related protein of the invention in a cell will
5 cause that cell to reduce its rate of poliferation. Similarly, sometimes decreasing expression of a cell proliferation-related protein of the invention in a cell will increase that cell's rate of poliferation; sometimes, this will cause that cell's rate of proliferation to decrease. What is relevant is that the rate of proliferation of the cell will change if the level of expression of a cell proliferation-related protein of the invention is either
10 increased or decreased.

Increasing the level of expression of a cell proliferation-related protein of the invention in a cell is a relatively simple matter. For example, overexpression of the protein can be accomplished by transforming the cell with a nucleic acid molecule encoding the protein according to standard methods such as those described above.

15 Reducing the level of expression of a cell proliferation-related protein of the invention in a cell is likewise simply accomplished using standard methods. For example, an antisense RNA or DNA oligonucleotide that is complementary to the sense strand (*i.e.*, the mRNA strand) of a nucleic acid molecule encoding the protein can be administered to the cell to reduce expression of that protein in that cell (see, *e.g.*,
20 Agrawal, U.S. Patent No. 5,929,226).

In another non-limiting example, RNAi can be employed to reduce the level of expression of a cell proliferation-related protein of the invention in a cell. RNAi (RNA interference) refers to the introduction of homologous double stranded RNA (dsRNA) to specifically target a gene's product, resulting in null or hypomorphic phenotypes. Thus,
25 because described herein are the nucleotide sequences encoding the cell proliferation-related proteins of the invention, RNAi can be readily designed. Indeed, constructs encoding an RNAi molecule have been developed which continuously synthesize an RNAi molecule, resulting in prolonged repression of gene expression of the targeted gene (Brummelkamp *et al.*, *Science* 296(5567): 550-3, 2002).

30 Protein expression levels can be measured by any standard method. For example, antibodies (monoclonal or polyclonal) can be generated by standard methods which

specifically bind to a cell proliferation-related protein of the invention (see methods for making antibodies in, *e.g.*, Ausubel *et al.*, *supra*; Current Protocols in Immunology, Coligan *et al.* (eds.), John Wiley & Sons, New York, NY, 1991, including updates up to 2002. Using such a cell proliferation-related protein-specific antibody, protein levels can
5 be determined by any immunological method including, without limitation, Western blotting analysis, immunoprecipitation, and ELISA.

Another non-limiting method for measuring protein level is by measuring mRNA levels. For example, total mRNA can be isolated from a cell introduced with a nucleic acid molecule of the invention (or with an antisense of such a nucleic acid molecule) and
10 from an untreated cell. Northern blotting analysis using as a probe the nucleic acid molecule which was introduced to the treated cell will readily demonstrate if the treated cell has a different level of expression of mRNA (and so a different level of expression of the encoded protein) as compared to the untreated cell.

Changes in cell proliferation rate (either in unchallenged cells and plants, or in
15 cells and plants challenged with, for example, exposure to salt or pathogen-infection) can be readily determined by counting the cells by any standard method. For example, cells can be manually counted using a hemacytometer or microscope. Callus growth and plant growth can be measured by weight and/or height. Individual cell growth can be determined by any standard cell proliferation assay (*e.g.*, ^3H incorporation).

20 The invention further includes manipulation of cell and plant proliferation by modulating the expression of more than one of the cell proliferation-related proteins described herein. For example, an increase in the level of expression of a first cell proliferation-related protein coupled with a decrease in the level of expression of a second the cell proliferation-related protein may result in a greater change in the cell
25 proliferation rate of a cell (or plant including such a cell) than either the increase in the level of expression of a first cell proliferation-related protein or the decrease in the level of expression of a second the cell proliferation-related protein alone. This invention has provided, in Figures 1-6 and the Examples below, numerous cell proliferation-related proteins and their interrelations with one another. Manipulation of expression of one or
30 more of the cell proliferation-related proteins of the invention enables the development of genetically engineered plants (*i.e.*, transgenic plants) that have superior growth rates

either in favorable conditions, under differentiation, or under stress (*e.g.*, biotic or abiotic stress).

The invention will be further described by reference to the following detailed examples. These examples are provided for purposes of illustration only, and are not
5 intended to be limiting unless otherwise specified.

Example I

Plant growth is accomplished two ways: by cell growth and by cell division, each of which is respectively controlled by the G1 phases and the M phases of the cell cycle.
10 Cyclins are proteins that play an active role in controlling nuclear cell division cycles, and regulate cyclin dependent kinases (CDKs), which are essential for cell cycle progression in eukaryotes. John *et al.* teaches that all cyclins interact with the catalytic subunit of cyclin-dependent protein kinases (CDK), and the two proteins (*i.e.*, the cyclin and CDK), along with the CDK activating subunit, in turn phosphorylate substrates on
15 serine or threonine residues, thereby controlling a chain of events that advance the cell through the various phases of the cell cycle (John *et al.*, *Protoplasma* 216(3-4): 119-142, 2001).

Eukaryotic cells have multiple classes of cyclins, each of which is required for specific regulatory steps during the cell cycle. Activity and substrate specificity of the
20 cyclin-CDK enzyme complex is determined by the specific cyclin subunit associated with the CDK catalytic subunit. Thus, the association of CDKs with specific cyclins is a key regulatory mechanism that advances the cell through the various stages of the cell cycle. Cell cycle progression involves changes in abundance of individual cyclins, due to changing rates of their transcription or proteolysis, with consequent changes in the
25 substrates of CDK through the cell cycle. Cyclin accumulation is particularly important in terminating the G1 phase, when such accumulation raises CDK activity and starts events leading to DNA replication.

Cyclins are essential for CDK activation and their binding to specific individual proteins is thought to provide potential substrates to CDKs (John *et al.*, *supra*). Thus, the
30 yeast two-hybrid approach was thought to be a useful method to dissect cyclin-mediated cell cycle events. Cyclin and CDK complex substrates include CDK inhibitors, kinases

and phosphatases, enzymes that control DNA replication, the cytoskeletal structures necessary for chromosome movement during mitosis, and compounds of the ubiquitin-dependent pathway for degradation of proteins, all of which participate in key steps of the cell cycle. High levels of CDK activity alternate with high levels of proteolytic activity, which is responsible for the turnover of cyclins and CDK inhibitors.

The eukaryotic cell cycle has a growth phase and a reproductive phase, the latter involving replication of chromosomes and their subsequent distribution to daughter cells. Cyclins are well conserved, and thus have been comparatively well characterized in plants. However, while the basic mechanisms of cell cycle control and the key genes that mediate cell cycle progression are highly conserved in eukaryotes (reviewed in Potuschak and Doerner, *Curr. Opin. Plant Biol.* 4(6): 501-506, 2001 and John *et al.*, *Protoplasma* 216(3-4): 119-42, 2001), some pathways regulating cell proliferation in plants are different from those in animals partly because plants are sessile and require developmental flexibility to respond to a spectrum of environmental changes (*e.g.*, flexible growth rates and patterns to exploit their environment optimally, cell division and expansion being essential to responding to environmental changes). Therefore, the pathways regulating cell proliferation in plants are likely different from those in animals. In higher plants, the cell cycle is coupled with developmental phase changes that are regulated by a complex gene network. (CDK-cyclin complexes and their involvement in cell cycle progression are reviewed by John *et al.*, *Protoplasma* 216:119-142, 2001). Plant cyclins and their associations with CDKs and substrate proteins are important and serve as key regulatory mechanisms that control proliferation in response to the many environmental and developmental cues that affect plant growth and development. (The role of cyclin-CDK complexes in regulation of the plant cell cycle is reviewed in John, *Protoplasma* 216:119-142, 2001 and Potuschak, *et al.*, *Curr. Opin. Plant Biol.* 4: 501-506, 2001).

This Example provides newly characterized rice proteins interacting with *O. sativa* E2F Homolog (OsE2F1) and identified by means of a yeast two-hybrid assay technology. One of the interactors found is a rice DP homolog similar to *Triticum sp.* DP Protein. This interactor was named Hypothetical Protein 018989-4003 (Os018989-4003) and was also used as a bait in the yeast two-hybrid screen.

In animals, members of the E2F transcription factor family regulate the expression of genes required for progression through the cell cycle, such as genes coding for several regulatory proteins and for enzymes involved in nucleotide and DNA synthesis. Specifically, E2F/DP complexes are important regulators of the G1/S transition (reviewed by Trimarchi and Lees, *Nat. Rev. Mol. Cell. Biol.* 3(1): 11-20, 2002), at which checkpoint cells either initiate the S phase or undergo arrest of the cell cycle. The E2F transcriptional activity results from the concerted action of a family of E2F-like proteins that form heterodimers. Based on sequence homology and functional properties of the genes that encode them, at least six E2F (E2F1 – E2F6) and two DP (DP1 and DP2) proteins have been identified in mammals as components of E2F complexes existing in all possible combinations. E2F subgroups (E2F1, E2F2 and E2F3, versus E2F4 and E2F5) are functionally different from each other and are thought to act in opposition to one another to mediate the activation or the repression of cell cycle regulator genes, thereby promoting either cellular proliferation or cell cycle arrest and terminal differentiation. Additionally, E2F activity is regulated by interactions with other cellular proteins including the three members of the retinoblastoma protein family Prb, p107 and p130, which bind to E2F and negatively regulate its transcriptional activity, and by indirect binding of cyclins and cyclin-dependent kinases (CDKs). Phosphorylation of Rb proteins by G1-specific CDKs releases the E2F heterodimer from the Rb protein in late G1 to S phase, and the resulting 'free E2F' induces the expression of many genes implicated in cellular proliferation, including cell cycle regulators and enzymes required for DNA synthesis. Individual E2F-DP complexes elicit different transcriptional responses depending on the identity of the E2F subunits and the proteins that are associated with the complex. These observations lend support to the yeast two-hybrid approach as a method to dissect E2F-mediated cell cycle control.

A number of cDNAs encoding E2F or DP homologs have been isolated from plants and characterized, including three E2F and two DP proteins from *Arabidopsis thaliana* (Magyar *et al.*, *FEBS Lett.* 486(1): 79-87, 2000; reviewed in Kosugi and Ohashi, *Plant Physiol.* 128(3): 833-843, 2002). Plant E2Fs share high sequence similarity but no distinguishable similarity with the animal E2F proteins, though they slightly resemble

E2F-4 and E2F-5. However, evidence is accumulating that plant E2F-like genes are functionally equivalent to their mammalian homologs and that the G1/S transition in plants is at least partly under the control of regulators similar to those found in animals, such as D-type cyclins, Rb-related proteins, and E2F and DP homologs. Like animal
5 E2Fs, plant E2F proteins can bind to the consensus binding sites of the animal E2F and their DNA-binding activities can be stimulated by human and plant DP proteins. They can also bind human Rb or plant Rb-like proteins. However, their properties, including transactivation, subcellular localization, and functional differences have not been well characterized (Kosugi and Ohashi, *supra*). One study indicates that, unlike animal E2Fs,
10 the *Arabidopsis* E2F and DP are not predominantly localized to the nucleus, but rather their nuclear localization is controlled by an interaction with some DPs or other proteins (Kosugi and Ohashi, *supra*). Based on these findings, Kosugi and Ohashi suggest that the function of plant E2F and DP proteins is primarily controlled by their nuclear localization mediated by the interaction with specific partner proteins, and that this
15 difference in the regulation of the E2F/Rb pathway between plants and animals may reflect differences in cell cycle regulation.

The protein interactions involving the rice E2F and DP homologs identified in this Example are aimed at elucidating the mechanisms of E2F-mediated cell cycle regulation in plants. Proteins that participate in cell cycle regulation in rice are targets for genetic
20 manipulation or for compounds that modify their level or activity, thereby modulating the plant cell cycle. The identification of genes encoding these proteins, as described herein, allows genetic manipulation of crops or application of compounds to modulate the plant cell cycle and effect agronomically desirable changes in plant development or growth.

25 Results

OsE2F1 was found to interact with four novel rice proteins: two DP-like proteins (Os018989-4003 and OsPN26539); a kinesin-like protein (OsPN29946) with a putative microtubule motor function in events occurring in the G1/S transition phase of the cell cycle; and a protein of unknown function (OsPN30852).

30 The novel DP protein Os018989-4003 (as either bait or prey in the yeast two-hybrid screen) interacted with rice E2F homolog OsE2F1 (described above) and with two

splicing variants of rice E2F2 homolog, OsE2F2 (annotated in the public domain) and OsE2F2 (367) (identified in this study). The OsE2F2 (367) variant also interacted with another novel DP-like protein, OsPN31182. Other interactors identified for the DP protein Os018989-4003 include rice kinesin-like protein (OsAAG13527); MADS box protein MADS14 (OsMADS14), with a known role in flower development; putative myosin heavy chain (OsAAK72891), which likely functions as an actin motor in cell-cycle-dependent cytoskeletal dynamic events; and another myosin heavy-chain-like protein, the novel protein OsPN22824.

The interacting proteins of this Example are listed in Tables 1 and 2 below, followed by detailed information on each protein and a discussion of the significance of the interactions. A diagram of the some of the interactions described in this Example is provided in Figure 1. The nucleotide sequences (from which the amino acid sequences can be deduced) of the proteins of this Example are provided in Figure 8.

Some of the proteins identified represent novel rice proteins previously uncharacterized. Based on their predicted biological function and on the ability of the prey proteins to specifically interact with rice E2F homolog OsE2F1 and DP homolog Os018989-4003, the interacting proteins are likely involved in the E2F-mediated regulation of the cell cycle.

Table 1. Interacting Proteins Identified for OsE2F1 (E2F Homolog).

The Myriad names and TMRI names of the clones of the proteins used as baits and found as preys protein name are given. Nucleotide/protein sequence accession numbers for the proteins of this Example (or related proteins) are shown in parentheses under the protein name. The bait and prey coordinates (Coord) are the amino acids encoded by the bait fragment(s) used in the search and by the interacting prey clone(s), respectively. The source is the library from which each prey clone was retrieved.

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (source)
BAIT PROTEIN:			
OsE2F1 PN19758	<i>O. sativa</i> E2F Homolog (AB041725; BAB20932)	300-437&	
INTERACTORS:			
Os018989-4003* PN21044	Hypothetical Protein 018989-4003, Similar to <i>Triticum sp.</i> DP Protein	100-250	9-179 177-294 (Output Trait)
OsPN26539	Novel Protein PN26539 (AC087544), Probable DP	100-250	2x 66-346 2x 194-346 82-253 (Output Trait)

OsPN29946	Novel Protein PN29946, Similar to <i>A. thaliana</i> Kinesin-Like Protein (BAB11329.1; e=0.0)	100-250	2x 173-470 (Output Trait)
OsPN30852	Novel Protein PN30852	100-250	45-86 (Output Trait)

& Self-activating clone, i.e., it activates the reporter genes in the two-hybrid system in the absence of a prey protein, and thus it was not used in the search

* This protein was also used as a bait in this Example (see Table 2).

5 **Table 2. Interacting Proteins Identified for Os018989-4003 (Hypothetical Protein 018989-4003, Similar to *Triticum* sp. DP Protein).**

Myriad /TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (Source)
BAIT PROTEIN :			
Os018989-4003 PN21044	Hypothetical Protein 018989-4003, Similar to <i>Triticum</i> sp. DP Protein		
INTERACTORS :			
OsE2F1 PN19758	<i>O. sativa</i> E2F Homolog (AB041725; BAB20932)	90-220	191-436 (Output Trait) 95-276 (Input Trait)
OsE2F2# PN21003	<i>O. sativa</i> E2F2 Homolog (AB041726; BAB20933)	90-220	90-358 (Input Trait)
OsAAG13527 PN23367	<i>O. sativa</i> Kinesin-like Protein (AC068924; AAG13527.1)	90-220	668-859 (Output Trait)
OsAAK72891 PN26317	<i>O. sativa</i> Putative Myosin Heavy Chain (AC091123; AAK72891)	90-220	342-638 322-549 (Input Trait) 339-651 (Output Trait)
OsMADS14* PN20910	<i>O. sativa</i> MADS Box Protein MADS14 (AF058697, AAF19047)	90-220	54-180 (Output Trait)
OsPN22824&	Novel Protein PN22824, Myosin heavy chain	90-220	2x 393-494 (Output Trait)

A splicing variant of the OsE2F2 sequence, OsE2F2 (367), was used as a bait; its interactions are shown below in Table 3

* Additional interactions identified for OsMADS14 are listed below on Table 4

& Additional interactions identified for PN22824 are listed below on Table 5

10 **Table 3. Interacting Proteins Identified for OsE2F2 (367)**

Myriad /TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (source)
BAIT PROTEIN :			
OsE2F2 (367) PN21003	E2F2 Homolog, Alt. Transcript (367) (AB041726; BAB20933)	180-368	

INTERACTORS:			
Os018989-4003	Hypothetical Protein 018989-4003, Similar to <i>Triticum sp.</i> DP Protein	1-368	69-294 (Input Trait)
OsPN31182	Novel Protein PN31183, <i>A. thaliana</i> DP-Like Protein (CAC15483.1; 9e ⁻⁵⁵)		124-324 72-255 156-334 (Input Trait)

Table 4. Additional interactions identified for OsMADS14:

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (source)
PREY PROTEIN:			
OsMADS14 PN20910	<i>O. sativa</i> MADS Box Protein MADS14 (AF058697, AAF19047)	50-198	124-223 82-197 (output trait)
BAIT PROTEIN:			
OsMADS45 PN20231 (1905929- OS000555)	<i>O. sativa</i> MADS Box Protein MADS45 (U31994, AAB50180)		

Table 5. Additional interactions identified for OsPN22824:

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (source)
PREY PROTEIN:			
OsPN22824	Novel Protein PN22824	1-198	301-500 (Input Trait)
BAIT PROTEIN:			
OsRACD PN19695	<i>O. sativa</i> Small GTP-Binding Protein RACDP (AF218381; AAF28764)		

Two-hybrid system using OsE2F1 as bait

OsE2F1 (GenBank Accession No. BAB209321; Kosugi and Ohashi, *Plant J.* 29(1): 45-59, 2002) is a 436-amino acid protein that is a member of the E2F transcription factor family. It contains a transcription factor E2F/dimerization partner (TDP) signature (amino acids 108 to 333), as predicted by analysis of the amino acid sequence (3.1e⁻³⁵ prediction value). E2F proteins function as heterodimers with transcription factors called DP proteins (Wu *et al.*, *Mol. Cell. Biol.* 15(5): 2536-2546, 1995). These transcriptional complexes regulate the transcription of genes encoding proteins required for progression through the cell cycle. Consistent with the interactions of E2F transcription factors with DP proteins documented in the literature are those identified in this Example between the rice orthologs of these proteins. It is likely that the Os018989-4003-OsE2F1 interaction

represents a step in cell cycle control in rice. This interaction was identified for both Os018989-4003 and OsE2F1 used as bait.

The bait fragment used in the yeast two-hybrid screen encoded amino acids 100 to 250 of OsE2F1.

5 OsE2F1 was found to interact with Os018989-4003, a protein of 294 amino acids that includes the presence of a transcription factor E2F/dimerization partner (TDP) signature (amino acids 100 to 294, $3.2e^{-17}$). E2F transcription factors form heterodimers with DP proteins; the resulting E2F/DP transcriptional complexes function as transcriptional activators of genes required for progression through the cell cycle (Wu *et al.*, *supra*). The activity of E2F/DP complexes is normally regulated by association with negative regulators of the retinoblastoma protein (pRB) family such as pRB, p107 and p130 and with other cellular proteins including cyclins and cyclin-dependent kinases (CDKs). Wu *et al.*, *supra*, also demonstrated that the binding specificity of the various E2F/DP complexes towards pRB or p107 is mediated by the E2F subunit. In agreement with the presence of the TDP signature, A BLAST analysis of the amino acid sequence of Os018989-4003 against the Genpept database indicated that this protein shares 62.5% identity with *Triticum sp.* DP protein (GenBank Accession No. CAC19034, 62.5%, e^{-91}). These analyses thus indicate that Os018989-4003 is a rice DP homolog.

20 Os018989-4003 was also used as a bait in the yeast two-hybrid screen. Its interactions are shown in Table 2 and discussed later in this Example.

25 OsE2F1 was also found to interact with novel protein OsPN26539. A BLAST analysis of the nucleotide sequence of the prey clone OsPN26539 identified the gene potentially encoding novel protein PN26539 on rice chromosome 10 clone nbxb0046P18A (GenBank Accession No. 26539). A BLAST analysis of the 346-amino acid sequence of OsPN26539 indicated that this protein is similar to putative protein (GenBank Accession No. NP_568116.1, 61% identity, $2e^{-103}$), Transcription Factor-Like Protein (GenBank Accession No. T48364, 56% identity, $6e^{-96}$), and DP-Like Protein (Accession # CAC15483, 53% identity, e^{-55}), all from *A. thaliana*. The DP-like protein is AtDPa, one of the two distinct DP-related proteins (AtDPa and AtDPb) identified in

Arabidopsis by Magyar *et al.*, *supra*. These authors showed that AtDPa and AtDPb heterodimerize *in vitro* with the *Arabidopsis* E2F-related proteins AtE2Fa and AtE2Fb identified by the same group. They also found that the AtDPa and AtE2Fa genes are transcribed in a cell cycle-dependent manner, being predominantly produced in actively
5 dividing cells, with highest transcript levels in early S phase cells. The novel protein OsPN26539 is thus likely a rice DP transcription factor.

OsE2F1 was also found to interact with novel protein OsPN29946. A BLAST analysis of the 614-amino acid sequence of OsPN29946 indicated that this protein is
10 similar to kinesin-like protein (GenBank Accession No. BAB11329.1, 70.9% identity, $e=0.0$) from *A. thaliana*. Kinesins are molecular motors, molecules that hydrolyze ATP and use the derived energy to generate motor force. Molecular motors are involved in diverse cellular functions such as vesicle and organelle transport, cytoskeleton dynamics, morphogenesis, polarized growth, cell movements, spindle formation, chromosome
15 movement, nuclear fusion, and signal transduction. Three families of non-plant molecular motors (kinesins, dyneins, and myosins) have been characterized. Kinesins and dyneins use microtubules, while myosins use actin filaments as tracks to transport materials intracellularly. A large number (about 40) of kinesin and myosin motors have been identified in *A. thaliana*, although little is known about plant molecular motors and
20 their roles in cell division, cell expansion, cytoplasmic streaming, cell-to-cell communication, membrane trafficking, and morphogenesis. Calcium, through the calcium binding protein calmodulin, is thought to play a key role in regulating the function of both microtubule- and actin-based motors in plants (molecular motors are reviewed in Reddy, A.S., *Int. Rev. Cytol.* **204**: 97-178, 2001). The kinesin-like
25 calmodulin (CaM) binding protein (KCBP), a minus end-directed microtubule motor protein unique to plants, has been implicated in cell division. During nuclear envelope breakdown and anaphase, activated KCBP promotes the formation of a converging bipolar spindle by sliding and bundling microtubules, while KCBP activity is down-regulated by Ca^{2+} and CaM during metaphase and telophase (Vos *et al.*, *Plant Cell* **12**(6):
30 979-990, 2000). The prey protein OsPN29946 is a kinesin-like protein likely involved in microtubule movements and its association with OsE2F1 suggests that this interaction

may represent a step in the control of cell-cycle dependent events involving cytoskeleton organization.

OsE2F1 was also found to interact with novel protein OsPN30852. A BLAST analysis of the 86-amino acid sequence of OsPN30852 indicated that this protein is similar to an unknown protein from *A. thaliana* (GenBank Accession No. AAK48957.1, 80% identity, $4e^{-31}$). Analysis of gene expression in plants indicated that this gene is up-regulated by stress and by abscisic acid and jasmonic acid.

Two-hybrid system using Os018989-4003 as bait

Hypothetical protein 018989-4003, which is similar to *Triticum sp.* DP Protein (Os018989-4003), was used as bait in the two-hybrid assay. This protein is described as an interactor for OsE2F1 earlier in this Example. The bait clone used in the screen encoded amino acids 90 to 220 of Os018989-4003.

The bait fragment encoding amino acids 90 to 220 of Os18989-4003 was found to interact with OsE2F1 (see description above). The interaction of Os018989-4003 with OsE2F1 confirms the interaction between the same proteins in the reverse bait and prey roles described earlier in this Example.

Os18989-4003 was also found to interact with OsE2F2. OsE2F2 is a protein of 393 amino acids that includes a transcription factor E2F/dimerization partner (TDP) (amino acids 74 to 300). A BLAST analysis indicated that this protein is the rice E2F homolog (GenBank Accession No. BAB20933, 100% identity, $e=0.0$), a member of the E2F transcription factor family. E2F transcription factor family members have been described herein. OsE2F2 is translated from one of two alternatively spliced mRNA species (identified in this study) and, like other E2F family members, it likely regulates transcription of genes encoding proteins involved in cell cycle progression in rice.

The splicing variant of OsE2F2, OsE2F2 (367), has a sequence of 367 amino acids that includes a predicted transcription factor E2F/dimerization partner (TDP) (amino acids 84 to 310, e^{-39} prediction value). A BLAST analysis of its amino acid sequence determined that it is the rice E2F homolog (GenBank Accession No.

5 BAB20933, 100% identity, $e=0.0$). OsE2F2 (367) was also used as a bait in this study and found to interact with following two DP proteins (these interactions are shown in Table 3):

- a) Hypothetical protein 018989-4003, which is similar to *Triticum sp.* DP Protein (Os018989-4003, described above). The OsE2F2 (367)-Os018989-4003
10 interaction validated the interaction between the same DP protein, namely 018989-4003, and OsE2F2.
- b) Protein PN31182, which is similar to *A. thaliana* DP-Like Protein (OsPN31182). OsPN31182 is a novel protein of 379 amino acids. A BLAST analysis indicated
15 that the amino acid sequence of OsPN31182 is similar to *A. thaliana* Putative Protein (top hit, GenBank Accession No. NP_568116.1, 70% identity, $5e^{-108}$) and DP-Like Protein (third hit, GenBank Accession No. CAC15483.1, 50% identity, $9e^{-55}$), and to DP-like proteins from other organisms. OsPN31182 is thus a novel rice DP protein.

20 DP proteins heterodimerize with E2F transcription factors to regulate the transcription of genes encoding proteins that are important for cell cycle progression. This notion is consistent with the interactions identified here between the rice E2F homolog OsE2F2 (367) and the DP-like proteins Os018989-4003 and OsPN31182. It is likely that these interactions participate in cell cycle progression in rice.

25 Os18989-4003 was also found to interact with OsAAG13527, an 859-amino acid protein determined by BLAST analysis to be the rice Kinesin-Like Protein (GenBank Accession No. AAG13527.1, 100% identity, $e=0.0$). Kinesins are molecular motors associated with microtubule movement during diverse cellular events, and have been
30 described herein.

Os18989-4003 was also found to interact with the putative myosin heavy chain protein OsAAK72891. A BLAST analysis of the OsAAK72891 amino acid sequence determined that this protein is the rice Putative Myosin Heavy Chain (GenBank Accession No. AAK72891.1, 100% identity, e=0.0).

5 Members of the myosin family participate in many types of cellular motility in all eukaryotic cells. Myosins are cytoskeletal proteins that function as molecular motors to generate movement and mechanical force in ATP-dependent interactions with actin filaments in various cellular events. The superfamily of myosin proteins has been divided into at least 14 classes (designated I to XIV) on the basis of their conserved
10 ATPase- and actin-binding regions, each myosin containing tail domains believed to be responsible for the specific subcellular localization and function of these motors (reviewed in Reichelt *et al.*, *Plant J.* **19**(5): 555-567, 1999). Molecular motors are involved in diverse cellular functions such as vesicle and organelle transport, cytoskeleton dynamics, morphogenesis, polarized growth, cell movements, spindle
15 formation, chromosome movement, nuclear fusion, and signal transduction (molecular motors are reviewed in Reddy, A.S., *Int. Rev. Cytol.* **204**: 97-178, 2001). While the role of myosins in animal and unicellular organisms is well established in muscular contraction, cytokinesis, and membrane-associated functions such as vesicle transport and membrane dynamics, little is known about myosins and other molecular motors in
20 plants and their roles in cell division, cell expansion, cytoplasmic streaming, cell-to-cell communication, membrane trafficking, and morphogenesis (Reddy, A.S., *supra*).

 Myosins in higher plants are thought to participate as motors in intracellular transport of organelles and vesicles associated with cytoplasmic streaming and in tip-growing cells of pollen tubes (reviewed in Yokota *et al.*, *Plant Physiol.* **121**(2): 525-534,
25 1999). The active sliding of myosin heavy chain along actin filaments provides the motor force for cytoplasmic streaming (*i.e.*, the constant movement of the cytoplasm and suspended organelles, membrane systems and molecules which is observed in plant cells), and the myosin activity is regulated by calcium through the calcium-binding protein calmodulin (Yokota *et al.*, *Plant Physiol.* **121**(2): 525-534, 1999; Yokota *et al.*,
30 *Plant Physiol.* **119**: 231-240, 1999). The function of cytoplasmic streaming and the mechanisms of its biochemical regulation are not known, although it is thought to

facilitate the exchange of materials within the cell and between the cell and its environment. Specific movement and anchoring of some organelles is also known to depend on actin filaments and is thus thought to involve myosin, but these mechanisms have not been documented (myosins are discussed in Biochemistry and Molecular
5 Biology of Plants, Buchanan, Gruissem and Jones (eds.), John Wiley & Sons, New York, NY 2002, p.221). Additionally, Reichelt *et al.* (*supra*) localized a plant myosin VIII at the post-cytokinetic cell wall, suggesting a role for this protein in cytokinesis, specifically in maturation of the cell plate and reestablishment of cytoplasmic actin cables at sites of intercellular communication. Based on current knowledge of plant myosins, the rice
10 heavy chain myosin OsAAK72891 may be a cytoskeletal component that participates in cytoplasmic streaming events in a cell-cycle-dependent manner.

Os18989-4003 was also found to interact with OsMADS14 (GenBank Accession No. AF058697), a 246-amino acid protein that includes a MADS box domain (amino
15 acids 1 to 61). Moon *et al.* report that OsMADS14 is homologous to the maize AP1 homolog ZAP1 and classify it as a member of the SQUAMOSA-like (SQUA) subfamily in the AP1/AGL9 family of MADS box genes, which control the specification of meristem and organ identity in developing flowers (Moon *et al.*, *Plant Physiol.* **120**(4): 1193-1204, 1999). OsMADS14 was expressed from the early through the later stages of
20 flower development, with transcripts detectable in sterile lemmas, paleas/lemmas, stamens, and carpels of mature flowers. Moon *et al.* suggested that this gene regulates a very early stage of flower development, based on their observation that transgenic plants ectopically expressing OsMADS14 exhibit extreme early flowering and dwarfism (Moon *et al.*, *supra*). MADS box proteins are known to regulate transcription as heterodimers or
25 ternary complexes that include other MADS box proteins, and these interactions are thought to occur through the K box present in MADS proteins (Lim *et al.*, *Plant Mol. Biol.* **44**(4): 513-527, 2000, Moon *et al.*, *supra*).

Because MADS box proteins are known to mediate various plant developmental processes as heterodimers or trimers, and given the involvement of the DP protein
30 Os018989-4003 in the regulation of genes required for cell cycle progression, it is likely the interaction between the MADS box protein OsMADS14 and Os018989-4003

represents a newly characterized interaction that regulates transcription of genes associated with plant development in rice.

OsMADS14 was also found to interact with the MADS box protein OsMADS45
5 (GenBank Accession No. AAB50180) (See Table 4). OsMADS45 is a 249-amino acid protein that includes a MADS box domain (amino acids 1 to 61) and two coiled coils (amino acids 83 to 117 and amino acids 152 to 176); the coiled coils are likely part of a K-box predicted between amino acids 73 and 176. The OsMADS45 gene, identified by Greco *et al.*, *Mol. Gen. Genet.* 253(5): 615-623, 1997, encodes a protein highly
10 homologous to the products of *Arabidopsis* AGL2 and AGL4 MADS box genes. Temporal and spatial RNA expression patterns suggest that the rice OsMADS45 and *Arabidopsis* AGL2 and AGL4 play similar roles in flower development (Greco *et al.*, *supra*), specifically in the development of all floral organs by acting as intermediates between the meristem identity and organ identity genes (Savidge *et al.*, *Plant Cell* 7(6):
15 721-33, 1995).

A BLAST analysis comparing the nucleotide sequence of OsMADS45 against TMRI's GeneChip[®] Rice Genome Array sequence database identified probeset OS014912_f_at ($6e^{-64}$ expectation value) and probeset OS000555_f_at ($6e^{-60}$) as the
20 closest matches. Analysis of gene indicated that these genes are expressed early in seed development.

Os18989-4003 was also found to interact with OsPN22824, a 500-amino acid protein fragment. A BLAST analysis of the OsPN22824 amino acid sequence revealed
25 no high similarity with any of the proteins in the Genpept database. The most similar amino acid sequences are six plant proteins of unknown function, the top hit being *A. thaliana* Expressed Protein (GenBank Accession No. NP_564015.1, 33% identity, $5e^{-45}$), and *A. thaliana* Myosin Heavy-Chain-Like (seventh hit, GenBank Accession No. BAA97502, 29% identity, e^{-016}). In agreement with these results, the most similar protein in Myriad's database is human Myosin, Heavy Chain IIX/d, Skeletal Muscle (MyHC-
30 IIX/d) (23% identity, $e=0.004$).

OsPN22824 was also found to interact with rice Small GTP-Binding Protein RACDP (OsRACD) (GenBank Accession No. AAF28764) (see Table 5). OsRACD is a 197-amino acid protein that includes an ATP/GTP-binding site motif A (P-loop, amino acids 13 to 20) and a prenyl group binding site (CAAX box, amino acids 194 to 197).

5 Analysis of the amino acid sequence by SMART identified a Rho (Ras homology) signature (amino acids 9 to 180, $6e^{-116}$), while analysis by Pfam predicted nearly the same region to be a Ras family signature (amino acids 8 to 197, $2.3e^{-78}$). These predictions indicate that OsRACD is a member of the Rho subfamily of Ras-like small GTPases. Hydrolysis of GTP to GDP is an important step in many intracellular signal transduction
10 pathways that control various cellular processes such as cell growth and development, apoptosis, lipid metabolism, cytoarchitecture, membrane trafficking, and transcriptional regulation (Aznar and Lacal, *Prog. Nucleic Acid Res. Mol. Biol.* 67: 193-234, 2001). The rice OsRACD protein has not been described, however, other members of the Rho subfamily have been characterized. Cdc42, Rac, and Rho isoforms regulate the assembly
15 and disassembly of the actin cytoskeleton in response to extracellular signals (Tapon and Hall, *Curr. Opin. Cell. Biol.* 9(1): 86-92, 1997). Plant small GTPase Rac homologs are components of the oxidative burst associated with disease resistance (Ono *et al.*, *Proc. Natl. Acad. Sci. USA* 98(2): 759-764, 2001; Dwyer *et al.*, *Biochim Biophys Acta* 1289(2): 231-237, 1996). OsRACD is a rice GTPase that likely participates in signal transduction
20 involving GTP hydrolysis, and its association with the myosin-like protein OsPN22824 suggests that this GTPase activity occurs during events related to organization of the actin cytoskeleton as part of either plant development and/or response to pathogen invasion.

25 Summary

OsE2F1 interacts with four novel rice proteins, two of which are DP-like proteins (Os018989-4003 and OsPN26539). In addition, the DP prey protein Os018989-4003 interacts with the E2F2 homolog splicing variant OsE2F2 (367) and, when used as bait, with both rice OsE2F1 and OsE2F2 homologs. OsE2F2 (367) also interacts with another
30 novel DP-like protein, OsPN31182. The identification of these new DP proteins interacting with E2F proteins in rice is in accord with the presence of E2F and DP

homologs identified previously in plants (reviewed in Kosugi and Ohashi, *Plant Physiol.* 128(3): 833-843, 2002). Plant E2F and DP proteins exhibit binding activities similar to those of animal E2F transcription factors, which function as heterodimeric complexes with DP or other E2F-like proteins (reviewed in Trimarchi and Lees, *Nat. Rev. Mol. Cell Biol.* 3(1): 11-20, 2002; Magyar *et al.*, *FEBS Lett.* 486(1): 79-87, 2000). The associations
5 between the rice E2F and DP homologs identified in this Example are consistent with the subunit composition of E2F/DP transcription factors and provide further evidence that plant E2F-like genes are functionally equivalent to their mammalian homologs. It is likely that these interactions participate in cell cycle progression in rice.

10

Animal E2F/DP transcription factors play a central role in the control of the G1/S transition through integration of the activities of important regulators of the cell cycle with the transcription apparatus. The G1/S control point in plants is thought to be at least partly regulated by molecules similar to those found in animals, such as D-type cyclins, Rb-related proteins, and E2F-like proteins (reviewed in Magyar *et al.*, *supra*). The G1
15 phase, which precedes the S phase, is a period of intense biochemical activity in which cells expand, double in size and synthesize molecules and structures, including microtubules and other cytoskeletal structures, in preparation for cell division. The end of G1 is an important checkpoint in the control of cell cycle progression, at which the control system either arrests the cycle or triggers initiation of the S phase (the plant cell
20 cycle phases are discussed in Biology of Plants, Raven, Evert and Eichhorn, 1999, Freeman/Worth, pp. 157-8). OsE2F1 and the DP protein Os018989-4003 were found to interact with several cytoskeletal structural proteins, and this finding supports the notion that the rice E2F/DP transcription factor has a role in controlling events related to cell
25 cycle progression. Two of these interactors are kinesin-like proteins: a novel rice kinesin-like protein (OsPN29946, interactor for OsE2F1) and rice kinesin-like protein annotated in the public domain (OsAAG13527, interactor for Os018989-4003). Two additional cytoskeletal components interacting with the DP protein Os018989-4003 are myosin heavy-chain proteins: putative myosin heavy chain (OsAAK72891) and a novel
30 rice myosin heavy-chain-like protein (OsPN22824). Kinesins and myosins are molecular motors that use microtubules (in the case of kinesins) or actin filaments (in the case of

myosins) as cytoskeletal tracks to transport cargo materials intracellularly. Molecular motors, including kinesins, myosins and dyneins, have been well characterized in non-plant organisms and implicated in a variety of cellular functions such as vesicle and organelle transport, cytoskeleton dynamics, morphogenesis, polarized growth, cell movements, spindle formation, chromosome movement, nuclear fusion, and signal transduction. In contrast, the roles of the many kinesins and myosins identified in plants are largely unknown (molecular motors are reviewed in Reddy, A.S., *supra*). A few studies suggest that myosin heavy-chain in higher plants participates in intracellular transport of organelles and vesicles (along actin filaments) associated with cytoplasmic streaming and in tip-growing cells of pollen tubes (reviewed in Yokota *et al.*, *Plant Physiol.* **121**(2): 525-534, 1999). An unconventional class VIII plant myosin has been implicated in maturation of the cell plate at cytokinesis (Reichelt *et al.*, *supra*). However, the function and regulation of plant motors in cell division, cell expansion, cytoplasmic streaming, cell-to-cell communication, membrane trafficking, and morphogenesis remains to be elucidated (Reddy, A.S., *supra*). Based on functional homology with animal and plant E2F proteins, which are known to participate in regulation of the G1/S transition phase, we speculate that the interactions of the rice OsE2F1 and DP protein Os018989-4003 with the kinesin-like and myosin-like prey proteins identified here represent transcriptional regulation of cell-cycle-dependent events involving cytoskeleton organization/function and possibly occurring during the G1/S transition.

Cell cycle regulators in plants must couple control of cell cycle phases to the environmental and developmental factors that affect plant growth and development. In agreement with this notion, the DP protein Os018989-4003 interacts with a protein known to regulate plant development, the MADS box protein MADS14 (OsMADS14), which in turn interacts with the MADS box protein OsMADS45. MADS box proteins mediate various plant developmental processes and, like other transcription factors, function as heterodimers or ternary complexes (for reviews, see Riechmann and Meyerowitz, *Biol. Chem.* **378**(10): 1079-1101, 1997; Moon *et al.*, *Plant Physiol.* **120**(4): 1193-204, 1999; Theissen *et al.*, *Plant Mol. Biol.* **42**(1): 115-149, 2000). (Interactions identified by our group for MADS box proteins are discussed below in Example IV).

The products of MADS box genes interact with each other and with other gene products participating in the genetic control of various plant development processes, with regulatory interactions (activation, repression) between the different genes/groups of genes within this network. Likewise, E2F-like proteins regulate transcription as heterodimeric complexes, and their activity is regulated by interactions with other cellular proteins (Trimarchi and Lees, *Nat. Rev. Mol. Cell. Biol.* 3(1): 11-20, 2002; Kosugi and Ohashi, *supra*). Given the presumed involvement of the DP protein Os018989-4003 in the regulation of genes required for cell cycle progression, it is likely that the interaction between the DP protein Os018989-4003, possibly in heterodimer form with OsE2F1 or OsE2F2 and the MADS box protein OsMADS14 is involved in transcriptional regulation of genes important in plant development in a cell-cycle dependent fashion in rice, and that these developmental processes may occur during the G1/S phase of the cell cycle.

The fourth interactor identified for E2F1 is a protein of unknown function (OsPN30852). However, based on its association with rice E2F1 and on the presumed role of the latter in regulation of cell cycle progression, it is likely that OsPN30852 is involved in cell cycle regulation.

The rice proteins found to interact with the rice E2F and DP homologs OsE2F1 and Os018989-4003 appear to be involved in regulation of the cell cycle/plant development. Some of these interactors are newly characterized rice proteins, and their interactions with OsE2F1 and Os018989-4003 represent molecular mechanisms for E2F-mediated transcriptional regulation of the cell cycle in rice that have not been previously described.

Example II

This Example provides newly characterized rice proteins interacting with rice cyclin OsS49462 and cyclin OsCYCOS2 identified by means of yeast two-hybrid assays.

As discussed in Example I, cyclins are regulatory proteins required to activate cyclin-dependent protein kinases (CDKs). Cyclins are classified into two groups: mitotic cyclins, which include A-type and B-type cyclins (also known as S and M cyclins,

respectively), which are essential for the control of the cell cycle at the G2/M (mitosis) transition, and G1 cyclins, which include D- and E-type cyclins, which are essential for the control of the cell cycle at the G1/S (start) transition. G2/M cyclins accumulate steadily during G2 and are abruptly destroyed as cells exit from mitosis (at the end of the M-phase).

B-type cyclins contain a large conserved central domain, the cyclin box, which interacts with the kinase subunit, and a domain called mitotic destruction box, which mediates cyclin degradation late in mitosis. B-type cyclins are expressed specifically in late G2 and early M phase of the cell cycle. They regulate the cell cycle progression from G2 to mitosis during plant development, and Myb-type transcription factors may be involved in this regulation (reviewed by Doonan, *et al.*, *Curr. Opin. Cell. Biol.* 9: 824-830, 1997). B-type cyclins of rice plants accumulate steadily during G2 and then are rapidly degraded at mitosis (Umeda, *et al.*, *Mol. Gen. Genet.* 262: 230-238, 1999). The B-type cyclins OsS49462 and OsCYCOS2 share 75.1% sequence identity at the amino acid level and are both encoded by mRNAs of 1.6 kb, as reported by Sauter *et al.*, *Plant J.* 7: 623-632, 1995. Expression of OsCYCOS2 is induced by the plant hormone gibberellin (GA) in the intercalary meristem of deepwater rice (*Oryza sativa L.*) internodes, and that the time course of OsCYCOS2 induction is compatible with a role for both cyclins in regulating the G2/M phase transition (Sauter *et al.*, *supra*). GA promotes rapid internodal growth in this plant subspecies, and this growth occurs through signaling events requiring cell cycle induction at the G2/M transition. Thus, GA promotes the activity of p34cdc2/CDC28-like histone H1 protein kinase, an enzyme known to regulate mitosis, and that the increase in this protein kinase activity is mediated by OSCYCOS2. The cyclins were expressed in the intercalary meristem and the elongation zone of the internode, but the GA-induced increase in transcript levels was restricted to the meristem only (Sauter *et al.*, *supra*).

Thus, OsS49462 and OsCYCOS2 are B-type mitotic cyclins that regulate the cell cycle progression from G2 to mitosis. The protein interactions involving OsS49462 and OsCYCOS2 identified in this Example are useful for elucidating the mechanisms of cell cycle regulation in plants. Proteins that participate in cell cycle regulation in rice may be targets for genetic manipulation or for compounds that modify their level or activity,

thereby modulating the plant cell cycle. The identification of genes encoding these proteins may allow genetic manipulation of crops or application of compounds to effect agronomically desirable changes in plant development or growth.

5 Results

Cyclin OsS49462 was found to interact with a rice hypothetical protein of unknown function (OsPN25358) and with four novel rice proteins: a putative RNA-binding protein (OsPN30848) and a zinc finger protein (OsPN29942), a myosin-like protein (OsPN23484) and an unknown protein (OsPN29957). Two of these proteins
10 (OsPN23484 and OsPN29942) also interact with the second bait, cyclin OsCYCOS2.

Cyclin OsCYCOS2 was found to interact with seven known rice proteins and with 18 novel rice proteins. The known interactors include a putative CCAAT displacement protein whose function as a transcriptional regulator is cell cycle-dependent (PN26210); a putative myosin heavy chain, a cytoskeletal protein that likely functions as a molecular
15 motor to move actin filaments in events related to cell polarity or cytokinesis (PN23297); a chloroplast ATPase I subunit (PN23416); a syntaxin related protein (PN23136); a heat shock protein (PN23169); a cora-like Mg transporter (PN25381) and a hypothetical protein of unknown function (PN23363). Among the novel interactors identified are several proteins with putative roles in cytoskeletal function: four putative myosin heavy-
20 chain proteins (PN23484, PN20815, OsPN29882, and OsPN29966); two kinesin-like proteins with a putative microtubule motor function during cell division (the calmodulin-binding protein OsPN23390 and the centromere/kinetochore protein OsPN29965); a spectrin-like protein with a presumed actin-binding function/nuclear matrix protein (OsPN29956); a putative Mg transporter (OsPN29970), a centromere homolog
25 (PN29958) and a zinc finger protein (PN29942). Other novel interactors include a protein similar to *A. thaliana* ARM repeat-containing protein with a possible role in cell adhesion and/or signaling (OsPN23274); a chaperone heat shock protein (PN30899); and 6 proteins of unknown function (OsPN29961, OsPN29969, OsPN26688, OsPN29967, OsPN29968, OsPN30854), two of which (OsPN23484 and OsPN29942) also interact
30 with the cyclin OsS49462 bait.

The interacting proteins of the Example are listed in Table 6 and Table 7 below, followed by detailed information on each protein and a discussion of the significance of the interactions. A diagram of the some of the interactions described in this Example is provided in Figure 1. The nucleotide and amino acid sequences of the proteins of this Example are provided in Figure 9.

Some of the proteins identified represent rice proteins previously uncharacterized. Based on their predicted biological function and on the ability of the prey proteins to specifically interact with cyclin OsS49462 and cyclin OsCYCOS2, the interacting proteins are likely part of a protein network involved in the cyclin-mediated regulation of the cell cycle.

Table 6. Interacting Proteins Identified for OsS49462 (Cyclin OsS49462, fragment).

The Myriad names and the TMRI names of the clones of the proteins used as baits and found as preys are given. Nucleotide/protein sequence accession numbers for the proteins of the Example (or related proteins) are shown in parentheses under the protein name. The bait and prey coordinates (Coord) are the amino acids encoded by the bait fragment(s) used in the search and by the interacting prey clone(s), respectively. The source is the library from which each prey clone was retrieved.

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (source)
BAIT PROTEIN :			
OsS49462 PN20325 (6331703-OS002997)	<i>O. sativa</i> Cyclin OsS49462, Fragment (X82035)	1-243 50-150 100-243	
INTERACTORS			
PN25358 13786464	Hypothetical Protein AAK39589	1 to 100	2x303-472 (output trait)
OsPN23484 Novel (CONTIG1447_FAST A.CONTIG1)	Novel Protein PN23484, heavy meromyosin	1 to 100	111-194 (output trait)
OsPN29942 novel	Novel Protein PN29942, Fragment, zinc finger protein	1 to 100	11-182 (output trait)
OsPN29957 novel	Novel Protein PN29957, Fragment, unknown	1 to 100	2x51-288 28-214 (output trait)
OsPN30848 novel	Novel Protein PN30848, Fragment, RNA binding protein	1 to 100	365-476 (input trait)

Table 7. Interacting Proteins Identified for OsCYCOS2 (*O. sativa* Cyclin OsCYCOS2).

The Myriad names and the TMRI names of the clones of the proteins used as baits and found as preys are given. Nucleotide/protein sequence accession numbers for the proteins of the Example (or related proteins) are shown in parentheses under the protein name. The bait and prey coordinates (Coord) are the amino acids encoded by the bait fragment(s) used in the search and by the interacting prey clone(s), respectively. The source is the library from which each prey clone was retrieved.

5

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (Source)
BAIT PROTEIN :			
OsCYCOS2 PN20257 (1694891-OS003088)	<i>O. sativa</i> Cyclin OsCYCOS2 (X82036)	1-150 100-275 140-350 300-420 1-420	
INTERACTORS:			
PN30899 417154	Hypothetical Protein 000221-3976 Similar to OsHP82, Fragment	50-233	4 to 228 (output trait)
PN29970	Putative CorA-like Mg ²⁺ Transporter Protein	50 to 233	1-158 (output trait)
PN23363 13324791	<i>O. sativa</i> Hypothetical Protein 13324791	50 to 233	50-148 (input trait)
PN26210 13702813	<i>O. sativa</i> Putative CCAAT Displacement Protein	170 to 310	422 to 646 2x364 to 613 (output trait)
15451591 PN23297	<i>O. sativa</i> Putative Myosin Heavy Chain	50 to 233	980 to 1160 (input trait)
PN23416 11466783	Chloroplast ATPase I Subunit	50 to 233	130 to 176 (input trait)
PN23136 5922624	Hypothetical Protein BAA85200 Similar to Syntaxin Related Protein AtVam3p	50 to 233	66 to 191 (output trait)
PN20815 Novel (3210-OS_ORF019753)	Hypothetical Protein PN20815 Similar to <i>A. thaliana</i> Myosin Heavy Chain, Fragment	170 to 310	1 to 134 (output trait)
OsPN23274 Novel (CONTIG697.FASTA. CONTIG2/ CONTIG697.FASTA. CONTIG1)	Novel Protein PN23274, Similar to <i>A. thaliana</i> ARM Repeat- Containing Protein	50 to 233	6x79 to 210 (input trait)
OsPN23390 novel	Novel Protein PN23390, Putative Kinesin-like Calmodulin Binding Protein, Fragment	50 to 233	595 to 845 576 to 738 (output trait)

OsPN23484 Novel (CONTIG1447.FAST A.CONTIG1)	Novel Protein PN23484, heavy meromyosin	170 to 310	77 to 233 2x64 to 212 90 to 245 (output trait)
OsPN26688 Novel (CONTIG3772.FAST A.CONTIG1)	Novel Protein PN26688, unknown	50 to 233	132 to 225 (input trait)
OsPN29882 novel	Novel Protein PN29882, Fragment, myosin heavy chain	50 to 233	107 to 273 (output trait)
OsPN29942 Novel (CONTIG3164.FAST A.CONTIG1)	Novel Protein PN29942, Fragment, zinc finger protein	170 to 310	1 to 159 (output trait)
OsPN29956 novel	Novel Protein PN29956, Fragment, nuclear matrix constituent	50 to 233	2x96 to 235 2 to 373 (output trait)
OsPN29958 novel	Novel Protein PN29958, Fragment, centromere homologue	50 to 233	3 to 304 (output trait)
OsPN29961 novel	Novel Protein PN29961, Fragment, Similar to <i>A. thaliana</i> Unknown Protein BAB02349	50 to 233	10 to 215 (output trait)
OsPN29965 novel	Novel Protein PN29965, Fragment, Similar to <i>A. thaliana</i> Kinesin (Centromere Protein)-Like Heavy Chain-Like Protein BAB03114	50 to 233	12 to 124 (output trait)
OsPN29966 novel	Novel Protein PN29966, Fragment, myosin heavy chain	50 to 233	8 to 216 (output trait)
OsPN29967 novel	Novel Protein PN29967, Fragment, unknown	50 to 233	3x16 to 174 (output trait)
OsPN29968 novel	Novel Protein PN29968, Similar to <i>A. thaliana</i> Unknown Protein BAB01990	50 to 233	12 to 113 (output trait)
OsPN29969 novel	Novel Protein PN29969, Similar to <i>A. thaliana</i> Unknown Protein BAB01990	50 to 233	2x16 to 123 (output trait)
OsPN25381 13357265	Protein 13357265 Putative CorA- like Mg ²⁺ Transporter Protein	50 to 233	30-218 (output trait)
OsPN30854 Novel (CONTIG962.FASTA. CONTIG1)	Novel Protein PN30854, unknown	170 to 310	100 to 169 (output trait)
OsPN30899 novel	Novel Protein PN30899, DNAJ	50 to 233	4 to 228 (output trait)

Two-hybrid system using OsS49462 as bait

The bait OsS49462 (GenBank Accession No. X82035; Sauter *et al.*, *Plant J.* 7 (4): 623-632, 1995) is a 242-amino acid protein that contains a cyclin, N-terminal domain (amino acids 1 to 105, $7.1e^{-49}$) and a cyclin C-terminal domain (amino acids 107 to 227, e^{-50}), as determined by analysis of the amino acid sequence. Like OsCYCOS2 (described as a bait below in this Example), OsS49462 is a rice B-type cyclin protein.

A BLAST analysis comparing the nucleotide sequence of OSS49462 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS002997.1_s_at ($e=0$ expectation value) as the closest match. Analysis of gene expression indicated that this gene is not specifically expressed in several different tissue types and is not specifically induced by a broad range of plant stresses, herbicides and applied hormones.

The bait protein encoding amino acids 1 to 100 of OsS49462 (which contains the cyclin, N-terminal domain) was found to interact with hypothetical protein AAK39589 (PN25358). Two prey clones encoding amino acids 303 to 472 of PN25358 were retrieved from the output trait library. PN25358 is a 472-amino acid protein that includes a transmembrane domain (amino acids 403 to 419), as predicted by analysis of the amino acid sequence. A BLAST analysis against the Genpept database determined that it is similar to a rice unknown protein (GenBank Accession No. AAK39589, $e=0$) and to an *A. thaliana* putative protein (GenBank Accession No. NP_199010.1, 64% identity, $7e^{-161}$). BLAST analysis of the PN25358 amino acid sequence against Myriad's proprietary database found no significant similarities for this protein. Since PN25358 interacts with OsS49462, it may be involved in cell cycle regulation.

The bait protein encoding amino acids 1 to 100 of OsS49462 was also found to interact with novel protein OsPN23484. One prey clone encoding amino acids 111 to 194 of OsPN23484 was retrieved from the output trait library. BLAST analysis suggests that PN23484 is a heavy meromyosin protein. Novel protein OsPN23484 also interacts with the bait OsCYCOS2 (described below in this Example). This observation validates the OsS49462-OsPN23484 interaction and suggests that OsPN23484 plays a broad role in regulation by cyclins and thus in the control of cell cycle progression.

The bait protein encoding amino acids 1 to 100 of OsS49462 was also found to interact with a fragment of the novel protein OsPN29942 (One prey clone encoding amino acids 11 to 182 of OsPN29942 was retrieved from the output trait library).

OsPN29942 is a protein for which the complete amino acid sequence is not known.

- 5 Analysis of the available 183 amino acids identified a BTB/POZ domain (amino acids 1 to 85). This domain is found primarily at the N terminus of zinc finger proteins and is evolutionarily conserved from *Drosophila* to mammals (Zollman, *et al.*, *Proc. Natl. Acad. Sci. USA* **91**: 10717-21, 1994). This region may affect the DNA-binding activity of zinc finger proteins (Bradwell, *et al.*, *Genes Dev.* **8**: 1664-1677, 1994). A BLAST
- 10 analysis against the Genpept database indicated that OsPN29942 shares 62% identity with an unknown protein from *A. thaliana* (GenBank Accession No. AAF00643, 5e⁻⁵³).

- OsPN29942 also interacts with the bait OsCYCOS2 as described later in this Example. This observation validates the OsS49462-OsPN29942 interaction and suggests that OsPN29942 plays a broad role in regulation by cyclins and thus in the control of cell
- 15 cycle progression.

- The bait protein encoding amino acids 1 to 100 of OsS49462 was also found to interact with OsPN29957. Three prey clones, two encoding amino acids 51 to 288 and one encoding amino acids 28 to 214 of OsPN29957 were retrieved from the output trait
- 20 library. OsPN29957 is a protein for which the complete amino acid sequence is not known. Upon analysis of the available 328 amino acids. A BLAST analysis against the Genpept database indicated that OsPN29957 shares 69% identity with an *A. thaliana* unknown protein (GenBank Accession No. NP_175186, e⁻²²). The available information makes it difficult to determine the function of OsPN29957. Discovery of the complete
- 25 amino acid sequence is likely to clarify the biological role of this protein and of its interaction with OsS49462.

- The bait protein encoding amino acids 1 to 100 of OsS49462 was also found to interact with PN30848 (One prey clone encoding amino acids 365 to 476 of OsPN30848 was retrieved from the input trait library). OsPN30848 is a protein for which the
- 30 complete amino acid sequence is not known. Analysis of the available 497 amino acids

identified two putative RNA-binding regions (amino acids 162 to 169 and amino acids 243 to 250). A BLAST analysis against the Genpept database indicated that OsPN30848 shares 50% identity with two *A. thaliana* putative RNA-binding proteins (GenBank Accession No. NP_190834, $2e^{-97}$ and GenBank Accession No. AAK32943, e^{-94}) and another *A. thaliana* protein similar to nucleolin (GenBank Accession No. AAB62861, 46% identity, $5e^{-89}$). Nucleolin is important for ribosome biogenesis and possesses RNA-binding activity. The similarity of OsPN30848 and nucleolin suggests a similar role for OsPN30848. The interaction of OsPN30848 with OsS49462 may alter cell cycle progression by regulating this activity.

10 A BLAST analysis comparing the nucleotide sequence of OsPN30848 against TMRI's GeneChip[®] Rice Genome Array sequence database identified probeset OS_ORF013388_at (e^{-108} expectation value) as the closest match. Gene expression analysis indicated that this gene is not specifically expressed in several different tissue types and is not specifically induced by a broad range of plant stresses, herbicides and
15 applied hormones.

Two-hybrid system using OsCYCOS2 as bait

The 419-amino acid protein OsCYCOS2 (GenBank Accession No. X82036; Sauter et al., Plant J. 7 (4): 623-632, 1995) is a G2/M type cyclin. Analysis of the
20 OsCYCOS2 amino acid sequence identified two cyclin domains spanning amino acids 200 to 284 ($2.7e^{-26}$) and amino acids 297 to 379 ($1.29e^{-22}$). Type G2/M cyclins regulate the cell cycle progression from G2 to mitosis during plant development. The role of these proteins has been discussed earlier in this Example with regard to the bait OsS49462.

25 A BLAST analysis comparing the nucleotide sequence of OsCYCOS2 against TMRI's GeneChip[®] Rice Genome Array sequence database identified probeset OS003088.1_at ($e=0$ expectation value) as the closest match. Gene expression analysis indicated that this gene is specifically expressed in panicle.

The bait encoding amino acids 50 to 233 of OsCYCOS2 was found to interact
30 with a fragment of the hypothetical protein 00221-3976 (PN30899). One prey clone

encoding amino acids 4 to 228 of PN30899 was retrieved from the input trait library. BLAST analysis indicates that PN30899 is most likely a heat shock (chaperone) protein (*Oryza sativa* protein 417154 HSP82). While heat shock proteins (HSPs) have been ascribed a main role in the plant stress response, some of these proteins are designated as
5 HSPs solely based on sequence homology and their functions in plants have not been demonstrated *in vitro*. Indeed, some HSPs are expressed throughout development. HSPs function as molecular chaperones that promote proper protein folding and may have roles not related to the stress response. HSP70 proteins, for instance, are essential for normal cell function. They are ATP-dependent molecular chaperones that may interact with
10 many different proteins, given their role in protein folding, unfolding, assembly, and disassembly. These topics are discussed in Biochemistry and Molecular Biology of Plants, Buchanan, Gruissem and Jones (eds.), John Wiley & Sons, New York, NY 2002. The heat shock protein HSP70 in sea urchin cells has been proposed to have a chaperone role in tubulin folding when localized on centrosomes, and in the assembling and
15 disassembling of the mitotic apparatus when localized on the fibres of spindles and asters (Agueli *et al.*, *Biochem. J.* 360(Pt 2): 413-419, 2001).

PN30899 also interacts with homeobox protein HOS59, fragment (OsHOS59) (see Example IV). Most proteins containing a homeobox domain are known to be sequence-specific DNA-binding transcription factors, some of which have important
20 roles in development. A BLAST analysis comparing the nucleotide sequence of PN30899 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS000221 _at (e=0 expectation value) as the closest match. Gene expression analysis indicated that this gene is not specifically expressed in several different tissue types and is not specifically induced by a broad range of plant stresses, herbicides and
25 applied hormones.

The bait encoding amino acids 50 to 233 of OsCYCOS2 was also found to interact with the putative Cor-A-like Mg^{2+} transporter protein, PN29970. (One prey clone encoding amino acids 1 to 158 of PN29970 was retrieved from the output trait
30 library.) The constitutively expressed CorA protein is the primary magnesium cation (Mg^{2+}) influx system of Bacteria and Archaea. CorA is ubiquitous in these organisms,

forming a distinct family of transport proteins that comprises at least 22 members, as determined by genomic sequence analysis, and with 6 more distant members in the yeasts (Kehres *et al.*, *Microb. Comp. Genomics* 3: 151-169, 1998). The similarity of PN29970 to a CorA protein suggests that this prey protein may function as an ion pump in events of the cell cycle regulated by OsCYCOS2.

The bait encoding amino acids 50 to 233 of OsCYCOS2 was also found to interact with hypothetical protein AAK18839 (PN23363) (GenBank Accession No. AC082645), a 286-amino acid protein in which no domains, motifs, or signatures have been clearly identified. (One prey clone encoding amino acids 50 to 148 of PN23363 was retrieved from the input trait library.) A BLAST analysis of the Genpept database indicates identity with an *O. sativa* unknown protein (GenBank Accession No. AAK18839, $3e^{-81}$). A BLAST analysis comparing the nucleotide sequence of PN23363 against TMRI's GeneChip[®] Rice Genome Array sequence database identified probeset OS_ORF005240_at (e^{-175} expectation value) as the closest match. Gene expression analysis indicated that this gene is not specifically expressed in several different tissue types and is not specifically induced by a broad range of plant stresses, herbicides and applied hormones.

A bait fragment encoding amino acids 170 to 310 of OsCYCOS2 was found to interact with the putative CCAAT displacement protein PN26210. Three prey clones, one encoding amino acids 422 to 646 and two encoding amino acids 364 to 613, of PN26210 were retrieved from the output trait library. PN26210 is a 687-amino acid protein that includes a transmembrane domain (amino acids 621 to 367), as predicted by analysis of the amino acid sequence. The analysis also predicted three coiled coils (amino acids 60 to 345, 381 to 445, and 489 to 643), although with prediction significance below threshold. Coiled coils participate in protein interactions in many types of proteins. A leucine zipper (amino acids 321 to 342) was also identified, which is known in transcription factors to facilitate dimer formation. Moreover, BLAST analysis of the amino acid sequence indicated that PN26210 is the same as *Oryza sativa* protein 13702813. CCAAT displacement proteins (known as CDP, Cut, or Cux in the literature)

belong to a highly conserved family of transcriptional regulators (reviewed by Nepveu, *Gene* 270: 1-15, 2001). These proteins have multiple DNA-binding domains that include one Cut homeodomain and one, two or three Cut repeats. The combination of these domains determines their distinct DNA-binding activities, which are elevated during proliferation and reduced during terminal differentiation. The CCAAT motif is found in the promoters of many eukaryotic genes, and CCAAT displacement proteins typically act as transcriptional repressors by directly binding to the promoters of genes that are important during development, but they can also function as transcriptional activators. CDP/Cut was found to be a component of the promoter complex HiNF-D, which is believed to promote the transcriptional induction of histone H4 genes at the G1/S phase transition of the cell cycle and to attenuate H4 gene transcription at later cell cycle stages in humans. The regulatory effect of CDP/Cut on transcription is thought to vary depending on the proteins with which it interacts (Nepveu, *supra*).

The bait encoding amino acids 50 to 233 of OsCYCOS2 was also found to interact with the putative myosin heavy chain protein PN23297. (One prey clone encoding amino acids 980 to 1160 of PN23297 was retrieved from the input trait library.) PN23297 (*Oryza sativa* protein 15451591) is a 1601-amino acid protein that includes an ATP/GTP-binding site motif A (P-loop) (amino acids 267 to 274). Analysis of the protein sequence clearly indicates that this protein is some form of myosin chain, being similar to many myosin-like proteins and myosin heavy chain proteins including myosin-like protein (GenBank Accession No. NP_195046, e=0.0) and myosin heavy chain (GenBank Accession No. T05200, e=0.0) from *A. thaliana*. While myosin is best known for its role in muscle contraction, this protein participates in other cellular events. In plants, for example, myosin heavy chain may participate in cytoplasmic streaming that occurs in tobacco and lily pollen tubes (Yokota *et al.*, *Plant Physiol.* 121: 525-534, 1999; Yokota *et al.*, *Plant Physiol.* 119: 231-240, 1999). Cruz *et al.* (*P. R. Health Sci. J.* 17: 323-326, 1998) present evidence that myosin assembly is important for mitosis. Specifically, myosin II-deficient yeast cells undergo cell cycle arrest at the G2/M transition, a phase regulated by OsCYCOS2. Furthermore, Xia *et al.* (*Plant J.* 10: 761-769, 1996) demonstrate that *A. thaliana* myosin heavy chain is among the proteins that

play a role in cell cycle regulation as well as in cytoskeleton function and in the establishment of cell polarity. The similarity of PN23297 to myosin heavy chain proteins suggests that this prey protein is a cytoskeletal component that may participate in events relating to cell polarity and cytokinesis.

5 Putative myosin heavy chain PN23297 also interacts with hypothetical protein 003118-3674 similar to *Lycopersicon esculentum* calmodulin (Os003118-3674). Os003118-3674 is a 148-amino acid protein with two EF-hand calcium-binding domains (amino acids 22 to 34 and 93 to 105). In agreement with the observation that Os003118-3674 includes EF-hand calcium-binding domains, BLAST analysis of the Genpept
10 database indicates that this protein shares 72% identity with *A. thaliana* putative calmodulin (GenBank Accession No. NP_1764705, e⁻⁵⁷), although the top score in this search is *A. thaliana* putative serine/threonine kinase (GenBank Accession No. NP_172695.1, 76% identity, 7e⁻⁶⁰). Therefore, this calmodulin-like protein may possess kinase activity. A BLAST analysis comparing the nucleotide sequence of putative
15 myosin heavy chain PN23297 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS005818_at (e⁻⁶ expectation value) as the closest match. The expectation value is too low for this probeset to be a reliable indicator of the gene expression of PN23297.

20 A bait fragment encoding amino acids 50 to 233 of OsCYCOS2 was also found to interact with the Chloroplast ATPase I subunit PN23415. One prey clone encoding amino acids 130 to 176 of PN23416 was retrieved from the input trait library. This protein shares the rice ATPase I subunit (GenBank Accession No. NP_039379; protein 11466783). ATPases are essential cellular energy converters that transduce the chemical
25 energy of ATP hydrolysis from transmembrane ionic electrochemical potential differences. The plant ATPases are present in chloroplasts, mitochondria and vacuoles. In the chloroplast, ATPases produce ATP that can be used as chemical energy in photosynthetic processes. The prey protein PN23416 is a chloroplast ATPase. A BLAST analysis comparing the nucleotide sequence of PN23416 against TMRI's GeneChip®
30 Rice Genome Array sequence database identified probeset OS003787_at (e=0 expectation value) as the closest match. Gene expression analysis that this gene is not

specifically expressed in several different tissue types and is not specifically induced by a broad range of plant stresses, herbicides and applied hormones.

5 A bait fragment encoding amino acids 50 to 233 of OsCYCOS2 was also found to interact with the hypothetical protein BAA85200 (*i.e.*, PN23136), which is similar to the syntaxin related protein AtVam3p. One prey clone encoding amino acids 66 to 191 of PN23136 was retrieved from the output trait library. PN23136 is *Oryza sativa* protein 5922624 (BAA85200) and is similar to AtVam3p. AtVam3p, the product of the AtVAM3 gene, is a syntaxin-related molecule implicated in vacuolar assembly in *A. thaliana*. This protein is expressed in various tissues including roots, leaves, inflorescence stems, flower buds, and young siliques, and AtVAM3 transcripts are abundant in undifferentiated cells in the meristematic region (Sato, *et al.* (1997) *J. Biol. Chem.* 272:24530-5). The AtVam3p protein is one of the t-SNARE membrane proteins that mediate protein cargo trafficking inside vesicles between the organelles of the plant endomembrane system. The AtVAM3p has been localized not only to the vacuolar membrane, but also on the prevacuolar compartment in *Arabidopsis* cells and has been suggested to also have a role in post-Golgi trafficking (Sanderfoot *et al.*, *Plant Physiol.* 121: 929-938, 1999). The similarity of PN23136 to a t-SNARE membrane protein and its association with OsCYCOS2 suggests that this prey protein may be involved in protein trafficking associated with the endomembrane system during the cell cycle.

25 A bait fragment encoding amino acids 170 to 310 of OsCYCOS2 was also found to interact with a fragment of the hypothetical protein PN20815, which is similar to the *A. thaliana* myosin heavy chain fragment. (One prey clone encoding amino acids 1 to 134 of PN20815 was retrieved from the output trait library.) PN20815 is a 496-amino acid protein. Analysis of the amino acid sequence determined that there is a possible cleavage site between amino acids 61 and 62, although no N-terminal signal peptide appears to be present. Its similarity to *A. thaliana* myosin heavy chain (GenBank Accession No. AAL11549, 4e⁻¹¹⁴) suggests that PN20815 might be a cytoskeletal component and may therefore participate in events relating to cell polarity and cytokinesis. Myosin assembly

is important for mitosis. Myosin proteins have been discussed herein with regard to the interacting protein PN23297.

A bait fragment encoding amino acids 50 to 233 of OsCYCOS2 was also found to
5 interact with novel protein PN23274. Six prey clones encoding amino acids 79 to 210 of
OsPN23274, a region that includes the putative leucine zipper in PN23274, were
retrieved from the input trait library. A BLAST analysis against the public databases
indicated that the 680-amino acid protein OsPN23274 is similar to *A. thaliana* putative
10 arm repeat containing protein (GenBank Accession No. NP_174228, e^{-80}) and to *Brassica*
napus putative arm repeat containing protein 1 (ARC1) (GenBank Accession No.
T08872, e^{-56}). Analysis of the OsPN23274 protein sequence predicted that it has an
armadillo/plakoglobin ARM repeat profile (amino acids 346 to 386; $1.8e^{-09}$). Two other
ARM-repeat domains were identified with much lower prediction significance (amino
acids 431 to 471, $e=1.2$; and amino acids 507 to 548, $e=35$). ARM motifs are tandemly
15 repeated sequences of approximately 50 amino acid residues that occur in a wide variety
of eukaryotic proteins (Peifer *et al.*, *Cell* **76**:789-791, 1994; Groves, *et al.*, *Curr. Opin.*
Struct. Biol. **9**: 383-389, 1999; Hatzfeld *Int. Rev. Cytol.* **186**: 179-224, 1999; Huber *et al.*
(1997) *Cell* **90**: 871-882, 1997). The ARM repeat was first identified in the *Drosophila*
protein armadillo that is involved in segment polarity and cell adhesion (Peifer *et al.*, *Cell*
20 **63**:1167-76, 1990). ARM repeats are found in the mammalian *Wnt* pathway proteins
beta-catenin (an armadillo homolog), plakoglobin, Adenomatous Polyposis Coli (APC)
tumor suppressor protein (Huber *et al.*, *supra*), and other proteins. The ARM repeats in
Armadillo family members mediate various protein interactions representing steps in
signaling events that result in control of cell adhesion, cytoskeletal alterations, and
25 transcription (reviewed by Hatzfeld, *supra*). Furthermore, analysis of the protein
sequence identified a SecD SecF domain (Bolhuis *et al.*, *J. Biol. Chem.* **273**: 21217-
21224, 1998) between amino acids 316 and 531, although with poor prediction
significance ($e=9$). This domain is necessary for secretion of some proteins. Also
predicted is a leucine zipper (amino acids 65 to 86), a domain known to facilitate protein
30 interactions, particularly in transcription factors. The predicted leucine zipper is of
interest when considering that beta-catenin is known to participate in transcriptional

regulation. Given its similarity to an ARM repeat protein and its interaction with OsCYCOS2, the prey protein OsPN23274 has a likely role in cell adhesion associated with cytoskeletal alterations occurring at the G2/M transition.

5 A BLAST analysis comparing the nucleotide sequence of OsPN23274 against TMRI's GeneChip[®] Rice Genome Array sequence database identified probeset OS017669 _at ($4e^{-70}$ expectation value) as the closest match. Gene expression analysis that this gene is not specifically expressed in several different tissue types and is not specifically induced by a broad range of plant stresses, herbicides and applied hormones.

10 A bait fragment encoding amino acids 50 to 233 of OsCYCOS2 was also found to interact with a fragment of the novel protein PN23390, a putative kinesin-like calmodulin-binding protein (OsPN23390). Two prey clones, encoding amino acids 595 to 845 and 576 to 738, of OsPN23390 were retrieved from the output trait library. Kinesins are molecular motors, molecules that hydrolyze ATP and use the derived energy
15 to generate motor force. Molecular motors are involved in diverse cellular functions such as vesicle and organelle transport, cytoskeleton dynamics, morphogenesis, polarized growth, cell movements, spindle formation, chromosome movement, nuclear fusion, and signal transduction. Three families of non-plant molecular motors (kinesins, dyneins, and myosins) have been characterized. Kinesins and dyneins use microtubules, while
20 myosins use actin filaments as tracks to transport materials intracellularly. A large number (about 40) of kinesin and myosin motors have been identified in *A. thaliana*, although little is known about plant molecular motors and their roles in cell division, cell expansion, cytoplasmic streaming, cell-to-cell communication, membrane trafficking, and morphogenesis. Calcium, through the calcium binding protein calmodulin, is thought
25 to play a key role in regulating the function of both microtubule- and actin-based motors in plants (molecular motors are reviewed in Reddy, *Int. Rev. Cytol.* 204: 97-178, 2001). The kinesin-like calmodulin (CaM) binding protein (KCBP), a minus end-directed microtubule motor protein unique to plants, has been implicated in cell division. During
30 nuclear envelope breakdown and anaphase, activated KCBP promotes the formation of a converging bipolar spindle by sliding and bundling microtubules, while KCBP activity is down-regulated by Ca^{2+} and CaM during metaphase and telophase (Vos *et al.*, *Plant Cell*

12: 979-990, 2000). The association of OsPN23390 with OsCYCOS2 suggests that the prey protein is involved in microtubule movement during cell division events mediated by the cyclin. The presence of a calmodulin-binding domain indicates that its activity is regulated by calmodulin.

5

OsCYCOS2 was also found to interact with the novel protein PN23484. The bait fragment used in the search encodes amino acids 170 to 310 of OsCYCOS2. Four prey clones, one encoding amino acids 77 to 233, two encoding amino acids 64 to 212, and one encoding amino acids 90 to 245, of OsPN23484 were retrieved from the output trait library. As already discussed above, OsPN23484 also interacts with the bait OsS49462. This observation validates the OsCYCOS2- OsPN23484 interaction and suggests that OsPN29942 plays a broad role in regulation by cyclins and thus in the control of cell cycle progression.

15

The bait fragment encoding amino acids 50 to 233 of OsCYCOS2 was also found to interact with novel protein OsPN26688. One prey clone encoding amino acids 132 to 255 of OsPN26688 was retrieved from the input trait library. OsPN26688 is a novel 251-amino acid protein of unknown function. The lack of information about OsPN26688 makes it difficult to determine its function and the significance of the OsCYCOS2-OsPN26688 interaction. However, the discovery of this interaction links OsPN26688 to control of the cell cycle in rice.

20

A BLAST analysis comparing the nucleotide sequence of OsPN26688 against TMRI's GeneChip[®] Rice Genome Array sequence database identified probeset OS005073.1 _at (e=0 expectation value) as the closest match. Gene expression analysis indicated that this gene is not specifically expressed in several different tissue types and is not specifically induced by a broad range of plant stresses, herbicides and applied hormones.

25

OsCYCOS2 was also found to interact with novel protein PN29882. This protein is similar to myosin proteins. The bait fragment used in the search encodes amino acids

30

50 to 233 of OsCYCOS2. One prey clone encoding amino acids 107 to 273 of OsPN29882 was retrieved from the output trait library.

OsPN29882 also interacts with MADS box-like protein BAA8188 (OsBAA81881) (see Example III). MADS box transcription factors, encoded by
5 members of the large MADS-box family of genes, participate in signal transduction and developmental control in plants, animals, yeast, and fungi. In plants, they are important regulators of genes implicated in flower and fruit development. This links cell cycling controlled by OsCYCOS2 to development controlled by MADS box proteins.

OsPN29882 also was found to interact with a ser/thr kinase/calmodulin that also
10 interacted with PN23297 (see description above). The ser/thr kinase/calmodulin may serve as part of the CDK complex with OsCYCOS2 to activate myosin substrates during mitosis.

A bait fragment encoding amino acids 170 to 310 of OsCYCOS2 (a region that
15 includes the cyclin domain) was found to interact with a fragment of the novel protein PN29942. This protein is discussed earlier in this Example as an interactor for the bait OsS49462. One prey clone encoding amino acids 1 to 159 of OsPN29942 was retrieved from the output trait library. This region spans the putative BTB/POZ domain that was identified in OsPN29942.

20 A bait fragment encoding amino acids 50-233 of OsCYCOS2 was found to interact with a fragment of the novel protein OsPN29956. OsPN29956 is a novel protein for which only a partial sequence is known. Analysis of the available 374 amino acids indicated that OsPN29956 includes a spectrin repeat (amino acids 167 to 209). In
25 agreement with the observations that OsPN29956 is a nuclear protein with a spectrin repeat, a BLAST analysis revealed that OsPN29956 shares amino acid sequence with nuclear matrix constituent protein 1 from *A. thaliana* (35% identity, Accession #BAB10684, $4e^{-55}$). Therefore, there is strong evidence that OsPN29956 is a nuclear matrix protein, and the interaction between OsCYCOS2 and OsPN29956 may represent a
30 step in cell cycle control through modulation of nuclear events.

Three prey clones were retrieved from the output trait library. Two of these encode amino acids 96 to 235 and one encodes amino acids 2 to 373 of OsPN29956. All three prey clones include the spectrin repeat that is present in OsPN29956. Spectrin repeats are also found in several proteins involved in cytoskeletal structure, such as actin-binding proteins (Hartwig, *Protein Profile* 2: 703-800, 1995). Actin-binding proteins of the superfamily of spectrins are ubiquitous proteins present in all animal and in plant cells. Spectrin-like epitopes have been localized mainly at the plasma membrane in several plant species and different cell types, but also in secretory vesicles, in the nuclei of various plant tissues, and in gravitropically tip-growing rhizoids and protonemata of characean algae, where they were found to be associated with the actin-organized aggregate of endoplasmic reticulum and correlated with active tip growth (Braun, *Plant Physiol.* 125:1611-1619, 2001). Studies indicate the presence of spectrin-based membrane skeleton in higher plant cells and demonstrate the ability of these proteins to interact with other components of the membrane skeleton such as actin and calmodulin (Bisikirska *et al.*, *Z. Naturforsch* 52: 180-186, 1997). Therefore, OsPN29956 could be a spectrin-like cytoskeleton protein that binds actin or calmodulin during events related to cell division.

A bait fragment encoding amino acids 50-233 of OsCYCOS2 was also found to interact with a fragment of protein PN29958. One prey clone encoding amino acids 3 to 304 of OsPN29958 was retrieved from the output trait library. BLAST analysis suggests that this is a centromere homologue (e-10) and is also homologous to the tobacco NT3 salinity tolerance protein (e-12). The BLAST results suggest a role for PN29958 in the centromere and also in salinity tolerance.

25

A bait fragment encoding amino acids 50-233 of OsCYCOS2 was also found to interact with protein PN29961, which is similar to *A. thaliana* protein BAB02349. One prey clone encoding amino acids 10 to 215 of OsPN29961 was retrieved from the output trait library.

30

A bait fragment encoding amino acids 50-233 of OsCYCOS2 was also found to interact with protein OsPN29965. (One prey clone encoding amino acids 12 to 124 of OsPN29965 was retrieved from the output trait library.) OsPN29965 is similar to *A. thaliana* kinesin (centromere protein). In animal cells, cytokinesis begins shortly after the sister chromatids move to the spindle poles. The centromere is a region of the chromosome to which the spindle fibers attach for the separation of the replicated chromatids in mitosis and meiosis. The kinetochores are the main sites of interaction between spindle microtubules and chromosomes; they are protein-rich structures associated with centromeric DNA and form on each sister chromatid at opposite sides of the paired centromeric region. Various proteins have been localized to animal kinetochores, including dynein and kinesin, but the protein composition of plant kinetocores has yet to be elucidated (Biochemistry and Molecular Biology of Plants, Buchanan, Gruissem and Jones (eds.), John Wiley & Sons, New York, NY 2002). The kinetochore-associated kinesin-like protein CENP-E binds to kinetochores during mitosis and has been shown to be essential for chromosome bioriented spindle attachment in mammalian cells (McEwen, *et al.*, *Mol. Biol. Cell.* **12**: 2776-89, 2001). Like CENP-E, the *Drosophila* kinesin-like motor protein CENP-meta similar to the vertebrate CENP-E, is a component of centromeric/kinetochore regions of *Drosophila* chromosomes and is required for maintenance of metaphase chromosome alignment (Yucel, *J. Cell. Biol.* **150**: 1-11, 2000). The inner centromere protein (INCENP) of animal cells has been implicated in both chromosome segregation and cytokinesis by promoting dissolution of sister chromatid cohesion and the assembly of the central spindle (Kaitna *et al.*, *Curr. Biol.* **10**:1172-1181, 2000). Kinesin-like proteins (KCBP) that are regulated by Ca^{2+} /calmodulin have been isolated from dicot (*A. thaliana*) as well as from monocot plants (maize). These motor proteins contain a highly conserved C-terminal region that includes the motor domain and the calmodulin-binding domain, which suggests that the KCBP is ubiquitous and highly conserved in all flowering plants (Abdel-Ghany, *et al.*, *DNA Cell Biol.* **19**: 567-578, 2000). Plant KCBP localizes to and is involved in establishing mitotic microtubule (MT) arrays during different stages of cell division, and Ca^{2+} /calmodulin regulates the formation of these MT arrays (Kao, *et al.*, *Biochem. Biophys. Res. Commun.* **267**: 201-207, 2000).

5

10

15

25

A bait fragment encoding amino acids 50-233 of OsCYCOS2 was also found to interact with a fragment of the novel protein PN29969, which is similar to the *A. thaliana* unknown protein BAB01990. Two prey clones encoding amino acids 16 to 123 of
5 OsPN29969 were retrieved from the output trait library. OsPN29969 is a novel protein for which the complete amino acid sequence is not known. Analysis of the available 123 amino acids identified a tropomyosin signature (amino acids 75 to 91), which suggests that OsPN29969 might be a novel structural protein. Tropomyosins are a family of
10 closely related proteins present in muscle and non-muscle cells. In striated muscle, tropomyosin mediates the interactions between the troponin complex and actin so as to regulate muscle contraction, while the role of this protein in smooth muscle and non-muscle tissues is not clear (Smilie, *Trends Biochem. Sci.* 4: 151-155, 1979; McLeod, *Bioessays* 6: 208-212, 1986). Based on the interaction of OsPN29969 with OsCYCOS2, this protein is likely to be involved in mediating interactions between actin and other
15 proteins during the G2/M transition. Thus, the interaction between OsCYCOS2 and OsPN29969 may represent a step in the control of the cell cycle through modulation of the nuclear matrix.

A bait fragment encoding amino acids 50-233 of OsCYCOS2 was also found to
20 interact with the putative Cor-A-like Mg^{2+} transporter protein PN25381. (One prey clone encoding amino acids 30 to 218 of OsPN25381 was retrieved from the output trait library.) This protein is *Oryza sativa* protein 13357265. The constitutively expressed CorA protein is the primary magnesium cation (Mg^{2+}) influx system of *Bacteria* and *Archaea*. CorA is ubiquitous in these organisms, forming a distinct family of transport
25 proteins that comprises at least 22 members, as determined by genomic sequence analysis, and with 6 more distant members in the yeasts (Kehres *et al.*, *Microb. Comp. Genomics* 3: 151-169, 1998). The similarity of PN25381 to a CorA protein suggests that this prey protein may function as an ion pump in events of the cell cycle regulated by OsCYCOS2.

30

A bait fragment encoding amino acids 170 to 310 of OsCYCOS2 was found to interact with novel protein PN30854. One prey clone encoding amino acids 100 to 169 of OsPN30854 was retrieved from the output trait library. OsPN30854 is a 169-amino acid protein. A BLAST analysis against the Genpept database indicated that OsPN30854
5 shares 67% identity with *A. thaliana* protein AT5g03660/F17C15_80 (GenBank Accession No. AAL06894, $9e^{-42}$). The interaction of PN30854 with OsCYCOS2 suggests that it plays some role in cell cycle regulation. A BLAST analysis comparing the nucleotide sequence of OsPN30854 against TMRI's GeneChip[®] Rice Genome Array sequence database identified probeset OS009560_r_at ($2e^{-16}$ expectation value) as the
10 closest match. The expectation value is too low for this probeset to be a reliable indicator of the gene expression of OsPN30854.

A bait fragment encoding amino acids 50 to 233 of OsCYCOS2 was found to interact with a fragment of novel protein PN30899, which is similar to *A. thaliana* protein
15 NP_199769. This protein is similar to DNAJ, a type of chaperone. Heatshock protein chaperones and potential roles in cell cycling have been discussed herein. One prey clone encoding amino acids 4 to 228 of OsPN30899 was retrieved from the output trait library.

20 Summary

M cyclins complexed with protein kinases commit the cell to mitosis at the G2-to-M transition. The synthesis of M cyclins in late G2 prepares the cell for mitosis, and increase of mitotic CDK activity at the G2-to-M transition initiates mitosis and cytokinesis. Mitosis, the stage in the cell cycle at which the duplicated chromosomes are
25 separated into two nuclei, and cytokinesis, the division of one cell into two cells, are accomplished by means of cytoskeletal structures. Mitosis depends on the mitotic spindle, a bipolar arrangement of mostly microtubules, but also actin and associated proteins, that interact with chromosomes and other proteins that participate in chromosome movement. Cytokinesis depends on the phragmoplast, an organelle
30 consisting of actin, myosin, and microtubules which gives rise to a plate in the center of the plant cell between the reforming nuclei and shapes the growing plate into a partition

in the form of a new cell wall. Actin filaments, microtubules, and intermediate filaments are filamentous protein polymers comprising the cytoskeleton of eukaryotic cells.

Accessory proteins are the motors and joints that link, move and modify the actin and tubulin scaffolding to stabilize the cytoskeleton, create polarities and move chromosomes during cell division, lower polymer concentration by binding (*i.e.*, proteins that bind soluble actin), and link the cytoskeleton to other cellular components such as biosynthetic or signaling enzymes. Many different accessory proteins mediate the function of the cytoskeleton by interacting with the polymers, including the motor proteins myosin, dynein and kinesin, as well as other proteins that cross-link (or bundle) cytoskeletal polymers of the same type. The dynamic behavior and polarity of actin and microtubules, enhanced by energy derived from hydrolysis of nucleoside triphosphates, is responsible for the movements of cytoplasm and organelles during the different phases of the cell cycle.

Mitosis starts with the initiation of chromosome condensation and the disassembly of the nuclear envelope that separates nuclear matrix from cytoplasm. Cells become fully competent for mitosis when the condensed chromosomes are aligned along a plane in the center of the cell, each chromosome comprising two chromatids (daughter strands) attached to each other and connected by microtubules to opposite ends of the cell. Chromosome segregation then initiates with the severing of the link between sister chromatids. The centromere is a region of the chromosome to which the spindle fibers attach for the separation of the replicated chromatids. The kinetochores, the main sites of interaction between spindle microtubules and chromosomes, are protein-rich structures that attach to centromeric DNA and serve as attachment points for the spindle microtubules, which congregate the chromosomes along a plate and subsequently pull apart the sister chromatids to opposite cell poles. Various proteins have been localized to animal kinetocores, including dynein and kinesin, but the protein composition of plant kinetocores has yet to be elucidated. (The plant cell cycle and cytoskeleton structure are discussed in detail in Biochemistry and Molecular Biology of Plants, Buchanan, Gruissem and Jones (eds.), John Wiley & Sons, New York, NY 2002). The concentrations of cyclins in the plant cell are thought to be important in mediating CDK

activity at the cytoskeleton, chromosomes, spindle, nuclear envelope, and phragmoplast (John *et al.*, *Protoplasma* 216(3-4): 119-142, 2001).

5 The interactions identified in this Example for OsCYCOS2 with several
cytoskeletal structural proteins in consistent with the role of the cyclin in controlling
events related to cell division. Five of these prey proteins—PN23484, PN23297,
PN20815, OsPN29882, and OsPN29966—are putative myosin heavy-chain proteins.
Previous reports on the role of *Arabidopsis* myosin heavy chain protein in cell cycle
control and cytoskeleton function Xia *et al.*, *Plant J.* 10(4): 761-769, 1996; Cruz, *et al.*, *P.*
10 *R. Health Sci. J.* 17: 323-326, 1998) suggest that the putative myosin prey proteins
identified here likely function as actin motors during the establishment of cell polarity at
mitosis or during cytokinesis. The observation by Cruz *et al.* that myosin is required in
yeast cells for the G2/M transition supports the notion that the interactions of
OsCYCOS2 with the myosin heavy chain proteins regulate the cell cycle at this transition
15 point. It is interesting that PN23297, PN29882 and PN29966 also interact with a ser/thr
kinase/calmodulin-like protein (Os003118-3674). Kinases regulate the activity of CDK-
cyclin complexes, and while no evidence exists that all three proteins—OsCYCOS2,
putative myosin heavy chain PN23297 (or other myosins), and the kinase Os003118-
3674—interact at the same time, the possibility that Os003118-3674 possesses kinase
20 activity increases the likelihood that this interaction propagates a signaling event.

Other cytoskeletal proteins interacting with OsCYCOS2 include a spectrin-like
protein with a presumed actin-binding function nuclear matrix constituent, and its
interaction with OsCYCOS2 may represent a step in cell cycle control through
25 modulation of nuclear events (OsPN29956).

Additional interactors with a motor function are the kinesin-like proteins
OsPN23390 and OsPN29965. Kinesins in both animals and plants are implicated in the
formation of mitotic spindles (Biochemistry and Molecular Biology of Plants, Buchanan,
30 Grissem and Jones (eds.), John Wiley & Sons, New York, NY 2002; Vos *et al.*, *Plant*
Cell 12: 979-990, 2000). Plant kinesin-like proteins regulated by calmodulin are

involved in microtubule array formation during cell division (Kao *et al.*, *Biochem. Biophys. Res. Commun.* **267**: 201-207, 2000). Based on these reports and on their interactions with OsCYCOS2, we postulate that the prey proteins OsPN23390 and OsPN29965 function as microtubule motor proteins during the formation of the mitotic spindle. The calmodulin-regulated OsPN23390 may be involved in microtubule array formation, while the similarity of OsPN29965 to a centromere protein suggests that this prey protein is a novel kinesin component of the centromeric/kinetochore regions of rice chromosomes with a putative role in chromosome alignment. The interactions of the cyclin protein with all these cytoskeletal proteins represent a newly characterized mechanism for control of cell division in rice.

OsCYCOS2 also interacts with PN23416, a protein similar to chloroplast ATPase I subunit. The interactions of the cyclin with microtubule- and actin-motor proteins is consistent with the presence of the ATPase prey protein. ATPases hydrolyze ATP to provide energy used by the motor proteins to generate force and directional movement associated with microtubules and actin filaments during mitosis.

Another prey protein, OsPN23274, is similar to *A. thaliana* ARM repeat-containing protein. The interactions of the ARM repeat domain with diverse binding partners reflect diverse functions for ARM repeat-containing proteins. These molecules combine structural roles as adhesion (cell-contact) and cytoskeleton-associated proteins with signaling roles by generating and transducing signals affecting gene expression (Hatzfeld, M., *Int. Rev. Cytol.* **186**: 179-224, 1999). The interaction of OsPN23274 with the cyclin suggests that the prey protein is likely involved in cell adhesion associated with the cytoskeletal alterations occurring during the transition from the G2 to M phase, although a role in signaling may be coupled with this function.

Another interactor for OsCYCOS2 is PN26210, a putative CCAAT displacement protein with a role as a transcriptional regulator. During replication, chromosomal DNA remains organized in chromatin, a complex composed mainly of histone proteins. Histone gene expression (RNA) and protein accumulation are strongly stimulated in early

S phase to double histone cellular content for the assembly of newly replicated DNA. CCAAT displacement proteins (CDPs) are thought to function as transcriptional activators of histone gene expression at the G1/S phase transition and as attenuators of histone gene transcription at later cell cycle stages in humans (Nepveu, A., *Gene* 270(1-2): 1-15, 2001). The dependence of the DNA-binding activity of these proteins on the cell cycle validates the interaction of a putative CCAAT displacement protein with a cyclin. Perhaps this interaction participates in a mechanism in which OsCYCOS2 sequesters PN26210 and prevents it from participating in gene regulation. It is also worth noting that the function of CDPs is regulated by posttranslational modifications (Nepveu, A., *supra*), specifically, the DNA-binding activity, and consequently, the transcriptional activity of CDP is inhibited by phosphorylation of either cut repeats or the cut homeodomain. Given that cyclins interact with cyclin-dependent kinases, it is tempting to speculate that the function of the OsCYCOS2-PN26210 interaction is, alternatively, to allow the posttranslational phosphorylation of PN26210 as part of the process leading to down-regulation of histone transcription during the G2/M phase.

Three membrane transport proteins were also found to interact with OsCYCOS2. PN23136 is similar to a t-SNARE membrane protein, a family of proteins involved in protein cargo trafficking among the organelles of the plant endomembrane system (Sanderfoot *et al.*, *Plant Physiol.* 121: 929-938, 1999). The ER system, which gives rise to the endomembrane system, is a dynamic network whose organization changes during the cell cycle. During mitosis, the ER undergoes a series of rearrangements that result in regulation of spindle activities and cell plate assembly through control of local calcium concentrations (Biochemistry and Molecular Biology of Plants, Buchanan, Gruissem and Jones (eds.), John Wiley & Sons, New York, NY 2002). The interaction of PN23136 with OsCYCOS2 points to a role for the prey protein in mediating protein trafficking associated with the dynamic behavior of the ER endomembrane system during mitosis. The other two transporters found to interact with OsCYCOS2 are putative CorA-like magnesium cation transporter which can function as a membrane-spanning pump to regulate turgor pressure or transmit solutes during cytokinesis.

Finally, OsCYCOS2 interacts with the putative heat shock prey proteins PN23169 and PN30899. HSPs act as molecular chaperones and, while these proteins in plants have been mainly linked to the stress response, some are not related to stress and their functions remain to be defined (Biochemistry and Molecular Biology of Plants,

5 Buchanan, Gruissem and Jones (eds.), John Wiley & Sons, New York, NY 2002). In the context of all the interactions identified for OsCYCOS2, we speculate that PN30899 and PN23169 act as a molecular glue to hold together interacting proteins. An alternative role for this prey protein may be deduced by functional homology with animal heat shock proteins whose chaperone roles in tubulin folding or mitotic structures
10 assembly/disassembly depends on their localization on centrosomes or spindle fibers, respectively (Agueli *et al.*, *Biochem. J.* 360(Pt 2): 413-419, 2001). These are functions associated with the phase of the cell cycle controlled by OsCYCOS2.

Proteins that participate in cell cycle regulation may be targets for genetic manipulation or for compounds that modify their level or activity, thereby modulating the
15 plant cell cycle. The identification of genes encoding these proteins in rice may allow the development of methods for controlling plant growth, specifically, cell proliferation and differentiation, to facilitate or retard plant development and promote regeneration. Such methods may involve the application of compounds to crops or the engineering of plants in which the level and/or activity of a protein associated with cell cycle regulation is
20 modulated for a time and under conditions sufficient to modify or control cell division.

One application for the results of this Example would involve modifying plant growth in the presence of one or more environmental conditions including increased or decreased temperatures, salinity, drought or nutrients, or exposure to disease. For example, in case that a limited amount of water is available following winter rain, it may
25 be necessary to restrain plant growth so that water resources are not exhausted before the valuable portion of the crop has developed. Chemical agents that reduce water transpiration have been found to have persisting adverse side effects on subsequent growth. By contrast, modulation of the expression or activity of proteins regulating the cell cycle could result in reduced growth without toxic side effects. Methods have been
30 proposed for controlling plant cell growth by modulating the level and or catalytic

activity of proteins having a cyclin-related kinase function to facilitate plant regeneration and development in cereal crops (see U.S. Patent Application No. 6087175 A1).

Example III

5 This Example provides a network of proteins interacting with rice MADS box protein MADS45 (OsMADS45), AP1-like MADS box protein (OsRAP1B), MADS box protein MADS6 (OsMADS6), MADS-box protein FDRMADS8 (OsFDRMADS8), MADS box protein MADS3 (OsMADS3), MADS box protein MADS5 (OsMADS5), and MADS box protein MADS15 (OsMADS15). Almost all the proteins of the network,
10 identified by means of yeast two-hybrid assays, are MADS box transcription factors.

 MADS box transcription factors, encoded by members of the large MADS-box family of genes, include a conserved sequence-specific DNA-binding/dimerization domain designated as the MADS box. These proteins participate in signal transduction
15 and developmental control in plants, animals, yeast, and fungi. In angiosperms, many MADS box proteins display primarily floral-specific expression and are important regulators of genes implicated in flower and fruit development, most notably in the determination of meristem and floral organ identity. Floral development is conserved among divergent species of flowering plants such as *Arabidopsis thaliana* and maize,
20 which indicates that MADS box genes are part of a highly conserved process that has evolved from an ancient flowering plant (the evolution and function of these genes is reviewed in Ng and Yanofsky, *Nat. Rev. Genet.* 2: 186-195, 2001), Thiessen *et al.*, *Plant Mol. Biol.*, 42:115-149, 2000), and specifically in rice and maize, in Munster *et al.*, *Gene* 262:1-130, 2001). Plant MADS box genes are organized into several phylogenetically
25 distinct gene groups—AGAMOUS (AG), APETALA3 (AP3)/PISTILLATA (PI) and APETALA1 (AP1)/ AG-LIKE (AGL)9—each group containing genes that share similar functions in regulating different aspects of flower development, including early acting meristem identity genes controlling the transition from vegetative to reproductive development and floral meristem development, late acting floral organ identity genes,
30 and genes mediating between these two functions (reviews by Purugganan *et al.*, *Genetics* 140: 345-356, 1995; Thiessen *et al.*, *Plant Mol. Biol.* 42: 115-149, 2000).

MADS box genes interact with each other and with other genes participating in the genetic control of flower development, with regulatory interactions (activation, repression) between the different genes/groups of genes within this network. In addition to flower development, several MADS box genes are involved in the control of ovule and seed development, vegetative growth, root development, fruit development and dehiscence, embryogenesis, or symbiotic induction (Moon *et al.*, *Plant Physiol.* 120: 1193-204, 1999; Riechmann and Meyerowitz, *Biol Chem.* 378: 1079-1101, 1997; Thiessen *et al.*, *supra*). Investigation of MADS box transcription factors and the proteins with which they interact in specific pathways can thus elucidate these biological processes at the molecular level.

The biological relevance of such interactions is further underlined by the fact that these proteins are known to regulate transcription as heterodimers or ternary complexes that include other MADS box proteins (Lim *et al.*, *Plant Mol. Biol.* 44:513-27, 2000). These interactions have been reported to occur through the K box (Sung *et al.*, *Mol. Cells* 11: 352-359, 2001; Lim *et al.*, *supra*) and to be enhanced by a region immediately downstream of the K domain. Plant MADS box proteins consist of a MADS box domain, an I region, a K domain, and a C-terminal region. The K box is a domain characteristic of plant MADS box proteins that sets them apart from their animal and fungal counterparts, which indicates that plant MADS box factors may have different criteria for interaction (Davies, *et al.*, *EMBO J.* 15:4330-4343, 1996). The K box is commonly found C-terminal to MADS box domains and is thought to serve as a dimerization moiety by forming coiled-coil structures known to facilitate protein interactions. The high potential for protein-protein interactions makes MADS box proteins suitable candidates for two-hybrid assays. However, though many MADS box proteins have been isolated from monocots including maize, sorghum, orchid and rice, few interactions between the MADS box proteins have been investigated (Moon *et al.*, *supra*). The protein interactions identified in this Example are aimed at elucidating the molecular mechanisms of plant development regulation by MADS box proteins in rice. The identification and characterization of protein interactions involving MADS box transcription factors in a major crop such as rice has important applications in agriculture.

Knowledge of the complex genetic system controlling flower morphogenesis in cereals could be exploited for the development of genetically engineered plants characterized as having a phenotype of modulated development, for example, early or delayed flowering.

5 A yeast two-hybrid search (as has been described above) led to the identification of a network of rice proteins comprised mainly of MADS box transcription factors that interact as heterodimers, some of which represent interactions not previously described. Some of the interactors are previously identified proteins including the MADS box proteins Os008339, OsFDRMADS6, OsMADS7, OsMADS8, OsMADS13, OsMADS14,
10 OsMADS17, OsMADS18, OsBAA81880, and the same proteins used as baits in these interaction studies, OsMADS45, OsRAP1B, OsMADS6, OsFDRMADS8. OsMADS1, OsMADS3, OsMADS5, and OsMADS15. An additional interactor is the seed storage protein prolamin (OsRP5). The search also led to the identification of six novel rice proteins: the MADS box protein OsPN29949 (interactor for OsMADS6); a putative
15 transcriptional regulator, OsPN23495 (interactor for OsMADS45); a putative hox protein, OsPN22834 (interactor for OsRAP1B); a protein of unknown function, OsPN31165 (interactor for OsMADS3); a 14-3-3-like protein, Os000564-1102 (interactor for OsMADS5); and a putative centromere protein, OsPN29971 (interactor for OsMADS15).

20 To determine the relationships among the interacting MADS box proteins, an analysis of the amino acid sequence alignment of the regions encoded by the interacting clones was performed. From these alignments, a phylogenetic tree was constructed.

25 The interacting proteins of the Example are listed in Tables 8-14, followed by detailed information on each protein and a discussion of the significance of the interactions. A diagram of the interactions is shown in Figure 2; the nucleotide and amino acid sequences of the proteins of this Example are provided in Figure 10. An analysis of the amino acid sequence alignments is shown in Figure 3A, and phylogenetic tree is shown in Figure 3B.

30

The ability of the interacting proteins to interact with the bait proteins OsMADS45, OsRAP1B, OsMADS6, OsFDRMADS8, OsMADS1, OsMADS3, OsMADS5, and OsMADS15, and the known or predicted biological functions of the interacting proteins indicate that the interacting proteins are involved in transcriptional regulation of genes associated with flower development in rice, except for prolamin, with a presumed role in seed development. Some of the interactions and proteins identified in this Example have not been previously described and represent a novel observation.

Tables 8-14. Interacting Proteins Identified in the Yeast Two-Hybrid Screen for the Bait Proteins OsMADS45, OsRAP1B, OsMADS6, OsFDRMADS8, OsMADS3, OsMADS5, and OsMADS15.

The Myriad names and the TMRI names of the clones of the proteins used as baits and found as preys are given. Nucleotide/protein sequence accession numbers for the proteins of the Example (or related proteins) are shown in parentheses under the protein name. The bait and prey coordinates (Coord) are the amino acids encoded by the bait fragment(s) used in the search and by the interacting prey clone(s), respectively. The source is the library from which each prey clone was retrieved.

Table 8. Interacting Proteins Identified for OsMADS45 (MADS box protein MADS45).

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (Source)
BAIT PROTEIN :			
OsMADS45 PN20231 (1905929-OS000555)	<i>O. sativa</i> MADS box protein MADS45 (U31994, AAB50180)	1-250* 100-250* 150-250*	
INTERACTORS :			
Os008339 PN20847(AJ293816-OS0083339)	<i>O. sativa</i> OS008339 MADS box transcription factor, fragment (AJ293816)	50-198	30-178 (input trait)
OsFDRMADS6 PN19766	<i>O. sativa</i> MADS-box protein FDRMADS6 (AF139664, AAF66997)	50-198	3x 115-246 93-244 (output trait)
OsFDRMADS8 PN20698	<i>O. sativa</i> MADS-box protein FDRMADS8 (AF141965, AAD38369)	50-198	2x 104-233 63-186 (output trait)
OsMADS1 PN19788 (11493806-OS015136)	<i>O. sativa</i> MADS box protein MADS1 (AF204063, AAG35652)	50-198	3x 82-241 2x 71-257 (output trait)
OsMADS3 PN20700	<i>O. sativa</i> MADS box protein MADS3 (L37528, AAA99964)	50-198	48-177 (output trait)
OsMADS5 PN20770	<i>O. sativa</i> MADS box protein MADS5 (U78890, AAB71434)	50-198	113-225 (output trait)

OsMADS6 PN20233	<i>O. sativa</i> MADS box protein MADS6 (U78782, AAB64250)	50-198	70-250 (output trait)
OsMADS13 PN20668	<i>O. sativa</i> MADS box protein MADS13 (AF151693, AAF13594)	50-198	2x 75-263 (output trait)
OsMADS14 PN20910	<i>O. sativa</i> MADS box protein MADS14 (AF058697, AAF19047)	50-198	124-223 82-197 (output trait)
OsMADS15 PN20842	<i>O. sativa</i> MADS box protein MADS15 (AF058698, AAF19048)	50-198	2x 92-237 (output trait)
OsMADS18 PN20912	<i>O. sativa</i> MADS box protein MADS18 (AF091458, AAF04972)	50-198	57-224 82-154 (output trait)
OsPN23495	Novel protein PN23495	50-198	39-165 12-198 (input trait)
OsRAP1B PN20232 (7592641- OS000556)	<i>O. sativa</i> AP1-like MADS box protein RAP1B (AB041020, BAA94342)	50-198	1-158 (output trait)

* Self-activating clone, i.e., it activates the reporter genes in the two-hybrid system in the absence of a prey protein, and thus it was not used in the search.

Table 9. Interacting Proteins Identified for OsRAP1B (*O. sativa* AP1-like MADS box protein RAP1B).

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (Source)
BAIT PROTEIN :			
OsRAP1B PN20232	<i>O. sativa</i> AP1-like MADS box protein RAP1B(AB041020, BAA94342)		
INTERACTORS:			
Os008339 PN20847	<i>O. sativa</i> OS008339 MADS box transcription factor, fragment (AJ293816)	1-150	3x 32-162 (input trait)
OsBAA81880 PN20837 (52957- OS011794)	<i>O. sativa</i> MADS box-like protein (AB003322, BAA81880)	125-235	2-168 24-203 (output trait)
OsFDRMADS6 PN19766	<i>O. sativa</i> MADS-box protein FDRMADS6 (AF139664, AAF66997)	1-247	1-186 (output trait)
		100-247	100-246 (output trait)
OsFDRMADS8 PN20698	<i>O. sativa</i> MADS-box protein FDRMADS8 (AF141965, AAD38369)	100-247	4x 69-233 (input trait) 94-230 (output trait)

		1-247	53-233 (output trait)
OsMADS1 PN19788	<i>O. sativa</i> MADS box protein MADS1 (AF204063, AAG35652)	1-247	4x 100-231 (input trait) 95-257 (output trait)
		100-247	2x 95-257 (input trait)
		65-200	4x 74-172 (input trait)
		125-235	73-239 (output trait)
OsMADS5 PN20770	<i>O. sativa</i> MADS box protein MADS5 (U78890, AAB71434)	30-180	106-225 (input trait) 121-225 (output trait)
		1-247	2x 109-225 (output trait)
		125-235	2x 108-225 (output trait)
OsMADS6 PN20233	<i>O. sativa</i> MADS box protein MADS6 (U78782, AAB64250)	1-247	116-250 (output trait)
OsMADS7 PN21116	<i>O. sativa</i> MADS box protein MADS7 (U78891, AAC49816)	1-247	5x 1-250 (output trait)
OsMADS8 PN20778	<i>O. sativa</i> MADS box protein MADS8 (U78892, AAC49817)	1-247	6x 107-248 (output trait) 75-248 (input trait)
		30-180	109-248 74-183 (output trait)
		100-247	127-248 (output trait)
		125-235	2x 79-248 (output trait)
OsMADS17 PN20914	<i>O. sativa</i> MADS box transcription factor MADS17 (AF109153, AAF21900)	1-247	106-249 (input trait)
OsMADS45 PN20231	<i>O. sativa O. sativa</i> MADS box protein MADS45 (U31994, AAB50180)	1-247	96-249 (input trait) 3x 75-249 (output trait)
		30-180	61-248 (output trait)
		125-235	4x 98-249 3x 69-249 (output trait)

OsPN22834	Novel protein PN22834, similar to Oshox6, fragment	1-247	2x 112-278 (input trait)
-----------	--	-------	-----------------------------

Table 10. Interacting Proteins Identified for OsMADS6 (*O. sativa* MADS box protein MADS6).

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (Source)
BAIT PROTEIN :			
OsMADS6 PN20233	<i>O. sativa</i> MADS box protein MADS6 (U78782, AAB64250)	1-251* 100-251*	
INTERACTORS :			
Os008339 PN20847	<i>O. sativa</i> OS008339 MADS box transcription factor, fragment (AJ293816)	50-200	108-226 (output trait)
OsBAA81880 PN20837	<i>O. sativa</i> MADS box-like protein (AB003322, BAA81880)	50-200	2x 120-228 (output trait)
OsFDRMADS8 PN20698	<i>O. sativa</i> MADS-box protein FDRMADS8 (AF141965, AAD38369)	50-200	91-233 (output trait)
OsMADS1 PN19788	<i>O. sativa</i> MADS box protein MADS1 (AF204063, AAG35652)	50-200	3x 70-257 (output trait)
OsMADS5 PN20770	<i>O. sativa</i> MADS box protein MADS5 (U78890, AAB71434)	50-200	61-171 (output trait)
OsMADS7 PN21116	<i>O. sativa</i> MADS box protein MADS7 (U78891, AAC49816)	50-200	95-259 (output trait)
OsMADS8 PN20778	<i>O. sativa</i> MADS box protein MADS8 (U78892, AAC49817)	50-200	2x 79-248 75-238 (output trait)
OsMADS15 PN20842	<i>O. sativa</i> OSMADS15 (AF058698, AAF19048)	50-200	73-183 1-176 (output trait)
OsMADS18 PN20912	<i>O. sativa</i> MADS box transcription factor MADS18 (AF091458, AAF04972)	50-200	64-249 (output trait)
OsMADS45 PN20231	<i>O. sativa O. sativa</i> MADS box protein MADS45 (U31994, AAB50180)	50-200	83-234 (output trait)
OsPN29949	Novel protein PN29949 putative MADS protein	50-200	118-241 109-193 (output trait)
OsRAP1B PN20232	<i>O. sativa</i> AP1-like MADS box protein RAP1B (AB041020, BAA94342)	50-200	1-188 (input trait)

			1-179 (output trait)
OsRP5 PN19877	<i>O. sativa</i> Prolamin (AF156714, AAF73991)	50-200	13-140 (output trait)

* Self-activating clone, i.e., it activates the reporter genes in the two-hybrid system in the absence of a prey protein, and thus it was not used in the search.

NOTE: Interactions of OsMADS6 with OsMADS14 and with OsMADS17, identified through a yeast two-hybrid system, are reported in the literature (Moon *et al.*, *Plant Physiol.* 120:1193-204, 1999).

5

Table 11. Interacting Proteins Identified for OsFDRMADS8 (*O. sativa* MADS box protein FDRMADS8).

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (Source)
BAIT PROTEIN :			
OsFDRMADS8 PN20698	<i>O. sativa</i> MADS-box protein FDRMADS8 (AF141965, AAD38369)		
INTERACTORS :			
OsMADS45 PN20231	<i>O. sativa</i> MADS box protein MADS45 (U31994, AAB50180)	60-160	3x 56-249 (output trait)

10

Table 12. Interacting Proteins Identified for OsMADS3 (*O. sativa* MADS box protein MADS3).

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (Source)
BAIT PROTEIN :			
OsMADS3 PN20700	<i>O. sativa</i> MADS box protein MADS3 (L37528, AAA99964)	120-210* 120-237*	
INTERACTORS :			
OsMADS8 PN20778	<i>O. sativa</i> MADS box protein MADS8 (U78892, AAC49817)	70-170	61-248 (input trait) 6-159 68-245 (output trait)
OsMADS45 PN20231	<i>O. sativa O. sativa</i> MADS box protein MADS45 (U31994, AAB50180)	70-170	48-249 (input trait) 4x 2-214 57-249 (output trait)
OsPN31165	Novel protein PN31165	70-170	58-252 (input trait)

* Self-activating clone, i.e., it activates the reporter genes in the two-hybrid system in the absence of a prey protein, and thus it was not used in the search.

15

Table 13. Interacting Proteins Identified for OsMADS5 (*O. sativa* MADS box protein MADS5).

Myriad/TMRI Gene	Protein Name	Bait Coord	Prey Coord
------------------	--------------	------------	------------

Name	(GenBank Accession No.)		(Source)
BAIT PROTEIN :			
OsMADS5 PN20770	<i>O. sativa</i> MADS box protein MADS5 (U78890, AAB71434)	100-226	
INTERACTORS :			
OsFDRMADS6 PN19766	<i>O. sativa</i> MADS-box protein FDRMADS6 (AF139664, AAF66997)	50-160	74-246 (output trait)
OsMADS13 PN20668	<i>O. sativa</i> MADS box protein MADS13 (AF151693, AAF13594)	50-160	2x 69-230 (output trait)
OsMADS17 PN20914	<i>O. sativa</i> MADS box transcription factor MADS17 (AF109153, AAF21900)	50-160	51-248 (output trait)
Os000564-1102 PN20072	Hypothetical protein 000564-1102	50-160	72-172 (output trait)
OsBAB56078 PN28517	<i>O. sativa</i> Hypothetical protein BAB56078 (AP003106, BAB56078)	50-160	51-155 (output trait)

Table 14. Interacting Proteins Identified for OsMADS15 (*O. sativa* MADS box protein MADS15).

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (Source)
BAIT PROTEIN :			
OsMADS15 PN20842	<i>O. sativa</i> MADS box protein MADS15 (AF058698, AAF19048)		
INTERACTORS :			
OsMADS1 PN19788 (11493806- OS015136	<i>O. sativa</i> MADS box protein MADS1 (AF204063, AAG35652)	100-235	95-254 4x 74-172 (input trait)
OsMADS45 PN20231	<i>O. sativa O. sativa</i> MADS box protein MADS45 (U31994, AAB50180)	100-235	120-249 (output trait)
OsPN29971	Novel protein PN29971, fragment, similar to <i>A. thaliana</i> centromere protein NP_191066	100-235	2x 1-108 (input trait)

5

O. sativa MADS box protein MADS45 (OsMADS45) as bait

OsMADS45 (GenBank Accession No. AAB50180; Greco *et al.*, *Mol. Gen. Genet.* 253(5): 615-623, 1997)) is a 249-amino acid protein that includes a MADS box domain (amino acids 1 to 61), as predicted by amino acid sequence analysis ($3.05e^{-41}$ prediction value). The analysis also predicted the existence of two coiled coils (amino acids 83 to 117 and amino acids 152 to 176). These coiled coils are likely part of a K-box predicted between amino acids 73 and 176 ($3.7e^{-45}$). The bait fragment used in this search encodes amino acids 50 to 198, a sequence that includes both predicted coiled coils and the K-box

10

of OsMADS45. OsMADS45 is highly homologous to the AGL2 and AGL4 MADS box genes, which are thought to play an important role in the development of all floral organs by acting as intermediates between the meristem identity and organ identity genes (Greco *et al.*, *Mol. Gen. Genet.* 253: 615-623, 1997; Savidge *et al.*, *Plant Cell* 7: 721-33, 1995).

- 5 In agreement with the expression pattern of AGL2 and AGL4, Northern blot and *in situ* hybridization experiments show that the rice OsMADS45 RNA is highly expressed in the floral meristem, in all the primordia, in mature floral organs, and in developing kernels (Greco *et al.*, *supra*), consistent with involvement in fruit development. However, temporal and spatial gene expression patterns only suggest that OsMADS45 and
- 10 *Arabidopsis* AGL2 and AGL4 play similar roles in flower development (Greco *et al.*, *supra*).

- A BLAST analysis comparing the nucleotide sequence of OsMADS45 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS014912_f_at ($6e^{-64}$ expectation value) and probeset OS000555_f_at ($6e^{-60}$) as the
- 15 closest matches. Analysis of gene expression indicated that these genes are expressed early in seed development.

- Proteins that were found to interact with OsMADS45 included Os008339 (GenBank Accession No. AJ293816), a 233-amino acid protein that includes a MADS
- 20 box domain (amino acids 10 to 67, $8.4e^{-29}$), which suggests that Os008339 is a member of the MADS box protein family. Analysis of the amino acid sequence also identified a K-box (amino acids 80 to 181) and a basic leucine zipper domain (bZIP) (amino acids 156 to 186). The bZIP domain is often found in transcription factors and includes a basic
- 25 DNA-binding region and a leucine zipper, which is associated with dimerization in many gene regulatory proteins (Landschulz *et al.*, *Science* 240: 1759-1764, 1988; Busch *et al.*, *Trends Genet.* 6: 36-40, 1990; O'Shea *et al.*, *Science* 243: 538-542, 1989). Thus this protein likely functions as do other MADS box family members, and its association with OsMADS45 represents a newly identified heterodimer presumably involved in
- 30 transcriptional regulation of genes associated with development in rice. The prey clone of Os008339 retrieved encodes a region that spans most of the K-box in Os008339. The

retrieval of this clone is consistent with OsMADS45 and Os008339 interacting through their respective K-boxes, as this domain is thought to include coiled coils used for protein interactions. Os008339 was also found to interact with the bait proteins OsRAP1B and OsMADS6 (see Table 9 and Table 10, respectively).

5 A BLAST analysis comparing the nucleotide sequence of Os008339 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS011977_i_at ($7e^{-91}$ expectation value) as the closest match. Gene expression analysis indicated that this gene is not specifically induced by a broad range of plant stresses, herbicides and applied hormones

10

OsMADS45 was also found to interact with *O. sativa* MADS box protein OsFDRMADS6 (GenBank Accession No. AF139664), a 246-amino acid protein that includes a MADS box domain (amino acids 1 to 61, $6.79e^{-39}$), a coiled coil located C-terminal to the MADS box domain (amino acids 116 to 182). This predicted coiled coil 15 is likely part of a K-box predicted between amino acids 73 and 174 ($8.9e^{-47}$), and its validity is supported by the fact that MADS box proteins bind DNA and modulate transcription as heterodimers. Previously published studies indicated that the FDRMADS6 transcript was present in flower, but not in root or shoot, and that transcripts were found in the spikelet apical meristem at the early stage of flower development and 20 again at the late stage when flower organ primordia began differentiating (Jia *et al.*, *Plant Sci.* 155:115-122, 2000). The OsFDRMADS6-OsMADS45 interaction has not been previously reported. OsFDRMADS6 was also found to interact with the bait proteins OsRAP1B (see Table 9) and OsMADS5 (see Table 13).

A BLAST analysis comparing the nucleotide sequence of OsFDRMADS6 against 25 TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS003005.1_i_at ($2e^{-82}$ expectation value) as the closest match. Gene expression analysis indicated this gene is not specifically induced by a broad range of plant stresses, herbicides and applied hormones.

30 OsMADS45 also interacted with OsFDRMADS8 (GenBank Accession No. AF141965), a 233-amino acid protein with a MADS box domain between amino acids 1

and 60 ($9.6e^{-39}$) and a coiled coil signature (amino acids 122 to 178, prediction significance below threshold), as determined by amino acid sequence analysis. This putative coiled coil region overlaps with a K-box domain (amino acids 73 to 173, $1.3e^{-10}$). While no information is available in the literature about OsFDRMADS8, the presence of the MADS box and the K-box strongly suggests that it is a transcription factor of the MADS box family. The association of this protein with OsMADS45 suggests a role for OsFDRMADS8 in transcriptional regulation of genes involved in plant development. The OsFDRMADS8- OsMADS45 interaction has not been previously reported. OsFDRMADS8 was also found to interact with the bait proteins OsRAP1B and OsMADS6 (see Table 9 and Table 10).

OsFDRMADS8 was also constructed as a bait. Its interactions are shown in Table 11 and described later in this Example. A BLAST analysis comparing the nucleotide sequence of OsFDRMADS8 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS015116 _at ($2e^{-82}$ expectation value) as the closest match. Analysis of gene expression indicated that this gene is not specifically induced by a broad range of plant stresses, herbicides and applied hormones.

The bait encoding amino acids 50 to 198 OsMADS45 was also found to interact with OsMADS1 (GenBank Accession No. AF204063), a 257-amino acid protein that is a member of the MADS box gene family. OsMADS1 includes a MADS domain (amino acids 1 to 60) and a coiled coil (amino acids 119 to 179), as determined by amino acid sequence analysis. OsMADS1 is a member of the AGL2 subfamily in the AP1/AGL9 family of MADS box genes (Moon *et al.*, *Plant Physiol.* 120(4): 1193-204, 1999). Ectopic expression of the OsMADS1 gene in homologous and heterologous plants results in early flowering, thereby suggesting a role for OsMADS1 in flower induction (Chung *et al.*, *Plant Mol. Biol.* 26(2): 657-665, 1994). OsMADS1 is expressed at the early stage through the later stages of flower development, with transcripts present in paleas/lemmas and carpels (Moon *et al.*, *supra*). The OsMADS1 homolog in the grass *Lolium temulentum* is expressed in the vegetative shoot apical meristem, and its expression increases strongly within 30 hours of long day floral induction, as determined by *in situ*

hybridization (Gocal *et al.*, *Plant Physiol.* 125(4): 1788-1801, 2001). The OsMADS1-OsMADS45 interaction has not been previously reported.

OsMADS1 was also found to interact with the bait proteins OsRAP1B (see Table 9), OsMADS6 (see Table 10), and OsMADS15 (see Table 14). A BLAST analysis
5 comparing the nucleotide sequence of OsMADS1 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS000262_f_at and OS015136_f_at ($5e^{-46}$ and $2e^{-36}$ expectation values, respectively) as the closest matches. Gene expression analysis indicated that this gene is not specifically induced by a broad range of plant stresses, herbicides and applied hormones.

10

OsMADS45 was also found to interact with the MADS box protein OsMADS3. The 236-amino acid OsMADS3 protein (GenBank Accession No. L37528), includes a MADS box domain (amino acids 1 to 61) and, based on sequence homology, is structurally and functionally related to the AG gene family, as reported by Kang *et al.*
15 (*Plant Mol. Biol.* 29:1-10, 1995). RNA blot analysis and *in situ* localization studies showed that the OsMADS3 RNA transcript is preferentially expressed in reproductive organs, especially in stamen and carpel. Transgenic plants engineered to ectopically express the OsMADS3 gene exhibit altered morphology and coloration of the perianth organs, suggesting an important role for OsMADS3 in flower development. The
20 OsMADS3-OsMADS45 interaction has not been previously reported.

OsMADS3 was also constructed as a bait protein. Its interactions are shown in Table 12 and described later in this Example. A BLAST analysis comparing the nucleotide sequence of OsMADS3 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS000554_f_at (e^{-43} expectation value) as the
25 closest match. Gene expression analysis indicated that this gene is not specifically induced by a broad range of plant stresses, herbicides and applied hormones.

OsMADS45 was also found to interact with the rice MADS box protein OsMADS5. OsMADS5 (GenBank Accession No. U78890) is a 225-amino acid protein
30 that includes a MADS box domain (amino acids 1 to 61, $3.17e^{-39}$), as predicted by amino acid sequence analysis. Thus, OsMADS5 is a member of the MADS box protein family.

Amino acid sequence analysis also predicted a coiled coil located C-terminal to the MADS box domain (amino acids 142 to 182), although with prediction significance below threshold. This coiled coil is likely part of a K-box predicted between amino acids 73 and 175 ($3.4e^{-40}$). OsMADS5 belongs to the AGL2 subfamily in the AP1/AGL9 family of MADS box genes, whose members are for the most part expressed at the early flowering stage (Moon *et al.*, *supra*). OsMADS5 is expressed throughout flower development, with higher expression in the early stages than the later stages and transcripts present in anthers and weakly in carpels, as reported by Kang *et al.* (*Mol. Cells* 7: 45-51, 1997). Transgenic plants ectopically expressing OsMADS5 exhibit the phenotype of weak dwarfism and early flowering, suggesting that this protein is involved in controlling flowering time. The OsMADS5- OsMADS45 interaction has not been previously reported.

OsMADS5 was also found to interact with the bait proteins OsRAP1B and OsMADS6 (see Table 9 and Table 10, respectively). OsMADS5 was also constructed as a bait protein. Its interactions are shown in Table 13 and described later in this Example. A BLAST analysis comparing the nucleotide sequence of OsMADS5 against TMRI's GeneChip[®] Rice Genome Array sequence database identified probeset OS011934 _at (e^{-58} expectation value) as the closest match. Analysis of temporal and spatial patterns of gene expression indicated that this gene is specifically expressed in panicle, in agreement with expression data previously reported for the OsMADS5 gene (Kang *et al.*, *supra*). Further, gene expression experiments indicated that the OsMADS5 gene is not specifically induced by a broad range of plant stresses, herbicides and applied hormones.

Os MADS45 was also found to interact with rice MADS box protein OsMADS6. OsMADS6 (GenBank Accession No. U78782) is a 250-amino acid protein that includes a MADS box domain (amino acids 1 to 59, $3.3e^{-42}$), as determined by amino acid sequence analysis. Thus, OsMADS6 is a member of the MADS box protein family. The analysis also predicted a K-box (amino acids 72 to 172, $3.4e^{-47}$). In support of the existence of a K-box, the analysis also predicted a coiled coil (amino acids 118 to 172). Moon *et al.* (*supra*) report that OsMADS6, like OsMADS14, belongs to the AP1/AGL9 family of genes which control the specification of meristem and organ identity in developing

flowers. Both OsMADS6 and OsMADS14 are expressed from the early through the later stages of flower development, with OsMADS6 transcripts detectable in lodicules and also weakly in sterile lemmas and carpels of mature flowers (Moon *et al.*, *supra*). Thus, these genes may regulate a very early stage of flower development, based on the observation
5 that transgenic plants ectopically expressing OsMADS6 and OsMADS14 exhibited extreme early flowering and dwarfism. The OsMADS6- OsMADS45 interaction has not been previously reported.

OsMADS6 was also found to interact with the bait protein OsRAP1B (see Table 9). OsMADS6 was also used as a bait. Its interactors are shown in Table 10 and
10 described later in in this Example. A BLAST analysis comparing the nucleotide sequence of OsMADS6 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS000571_f_at (e^{-7} expectation value) as the closest match. The expectation value is too low for this probeset to be a reliable indicator of the gene expression of OsMADS6.

OsMADS45 was also found to interact with rice MADS box protein OsMADS
15 13). OsMADS13 (GenBank Accession No. AF151693) is a 250-amino acid protein that includes a MADS box domain (amino acids 1 to 61). Lopez-Dee *et al.* (*Dev. Genet.* 25: 237-244, 1999) determined that this gene is the ortholog of ZAG2, a maize MADS-box
20 gene expressed mainly in the ovule, and of the ZAG2 paralogous gene ZMM1. The OsMADS13 gene is highly expressed in developing ovules and may play a role in rice ovule and seed development (Lopez-Dee *et al.* *supra*). Ovules are contained in the carpel, structures in the flowers of seed plants such as rice, and they develop into seeds after fertilization. The OsMADS13- OsMADS45 interaction has not been previously
25 reported.

OSMADS13 was also found to interact with the bait protein OSMADS5 (see Table 13). A BLAST analysis comparing the nucleotide sequence of OsMADS13 against
30 TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS000554_f_at (e^{-77} expectation value) as the closest match. Gene expression analysis

indicated that this gene is not specifically induced by a broad range of plant stresses, herbicides and applied hormones.

OsMADS45 was also found to interact with rice MADS box protein OsMADS14.
5 OsMADS14 (GenBank Accession No. AF058697) is a 246-amino acid protein that includes a MADS box domain (amino acids 1 to 61). OsMADS14 is homologous to the maize AP1 homolog ZAP1 and is a member of the SQUAMOSA-like (SQUA) subfamily in the AP1/AGL9 family of MADS box genes, which control the specification of meristem and organ identity in developing flowers (Moon *et al.*, *supra*). OsMADS14,
10 as well as OsMADS6, is expressed from the early through the later stages of flower development, with OsMADS14 transcripts detectable in sterile lemmas, paleas/lemmas, stamens, and carpels of mature flowers. Thus, these genes may regulate a very early stage of flower development, based on the observation that transgenic plants ectopically expressing OsMADS14 and OsMADS6 exhibit extreme early flowering and dwarfism
15 (Moon *et al.*, *supra*). The OsMADS14-OsMADS45 interaction has not been previously reported.

OsMADS14 was also found to interact with Os018989-4003 (hypothetical protein 018989-4003 similar to *Triticum sp.* DP Protein). Using a yeast two-hybrid system,
20 OsMADS14 has also been reported to interact with OsMADS1 (Lim *et al.*, *Plant Physiol.* 120: 1193-1204, 1999) and with OsMADS6 (Moon *et al.*, *supra*). While the K domain is essential for the interaction between OsMADS14 and OsMADS1, a region preceded by the K domain augments this interaction (Lim *et al.*, *supra*). Likewise, a 14-amino acid region located immediately downstream of the K domain enhances the
25 OsMADS14-OsMADS6 interaction, and the two leucine residues within this region play an important role in that enhancement (Moon *et al.*, *supra*). A BLAST analysis comparing the nucleotide sequence of OsMADS13 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS003005.1_i_at (e^{-82} expectation value) as the closest match. Gene expression analysis indicated that this gene is not
30 specifically induced by a broad range of plant stresses, herbicides and applied hormones.

OsMADS45 was also found to interact with rice MADS box protein OsMADS 15. OsMADS15 (GenBank Accession No. U78782) is a 267-amino acid protein with a MADS box domain between amino acids 1 and 60, as determined by amino acid sequence analysis ($5.39e^{-42}$ prediction value). The analysis also predicted a coiled coil signature (amino acids 145 to 184). This putative coiled coil region overlaps with a predicted K-box domain (amino acids 73 to 174, $1.20e^{-40}$). OsMADS15 is homologous to the maize AP1 homolog ZAP1 and is classified as a member of the SQUAMOSA-like (SQUA) subfamily in the AP1/AGL9 family of MADS box genes, which control the specification of meristem and organ identity in developing flowers (Moon *et al.*, *supra*). The OsMADS15- OsMADS45 interaction represents a heterodimer that has not been previously reported.

OsMADS15 was also found to interact with the bait protein OsMADS6 (see Table 10). OsMADS15 was also constructed as a bait protein. Its interactions are shown in Table 14 and described later in this Example. A BLAST analysis comparing the nucleotide sequence of OsMADS15 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS015053_f_at (e^{-77} expectation value) as the closest match. Gene expression analysis indicated that this gene is not specifically induced by a broad range of plant stresses, herbicides and applied hormones.

OsMADS45 was also found to interact with rice MADS box protein OsMADS18. OsMADS18 (GenBank Accession No. AF091458) is a 249-amino acid protein with a MADS box domain between amino acids 1 and 60 ($1.67e^{-38}$), as determined by amino acid sequence analysis. This amino acid sequence analysis also predicted a coiled coil signature (amino acids 141 to 191). This putative coiled coil region overlaps with a K-box domain (amino acids 73 to 173, $3.80e^{-32}$). OsMADS18 is highly homologous to the maize AP1 homolog ZAP1 and belongs to the SQUA subfamily in the AP1/AGL9 family of MADS box genes, which control the specification of meristem and organ identity in developing flowers (Moon *et al.*, *supra*). The OsMADS18- OsMADS45 interaction represents a heterodimer that has not been previously reported.

OsMADS18 was also found to interact with OsMADS6 (see Table 10). A BLAST analysis comparing the nucleotide sequence of OsMADS18 against TMRI's

GeneChip® Rice Genome Array sequence database identified probeset OS015196_i_at (e^{-58} expectation value) as the closest match. Gene expression analysis indicated that this gene is not specifically induced by a broad range of plant stresses, herbicides and applied hormones.

5

OsMADS45 was also found to interact with the novel rice protein OsPN23495. OsPN23495 is a novel 335-amino acid protein. A BLAST analysis indicated that OsPN23495 is similar to expressed protein from *A. thaliana* (GenBank Accession No. NM_129661, 42.1% identity, $2e^{-054}$), for which no information is available in the public domain. However, OsPN23495 was also found to interact with two rice hypothetical proteins (Os006111-3329 and Os020134-3170) which are similar to the zinc/DNA-binding ascorbate oxidase promoter binding protein (AOBP) from *Curcubita maxima*, and which include a Dof domain zinc finger DNA-binding domain (amino acids 103 to 165, $1.9e^{-37}$ for Os006111-33229; amino acids 101 to 163, $3.8e^{-38}$ for Os020134-3170). The presence of the Dof domain suggests that these two proteins are transcriptional regulators. Thus, by virtue of its interaction with these two proteins and with OsMADS45, novel protein PN23495 may be a novel transcription factor involved in regulation of genes controlling plant development. The OsPN23495-OsMADS45 interaction is a newly identified interaction.

20 A BLAST analysis comparing the nucleotide sequence of OsPN23495 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS001986_at ($e=0$ expectation value) as the closest match. Gene expression analysis indicated that this gene is not specifically induced by a broad range of plant stresses, herbicides and applied hormones.

25

OsMADS45 was also found to interact with AP-1 like MADS box protein OsRAP1B. OsRAP1B (GenBank Accession No. AB041020) is a 246-amino acid protein encoded by a member of the MADS box gene family. It includes a MADS box domain between amino acids 1 and 60. OsRAP1B was identified by Kyoizuka *et al.* (*Plant Cell Physiol.* 41:710-718, 2000) as a putative rice ortholog of the *Arabidopsis* APETALA1

30

(AP1), a class of MADS box genes involved in specification of floral organ identity. The OsRAP1B-OsMADS45 interaction has not been previously reported.

OsRAP1B was also constructed as a bait. Its interactors are listed in Table 9 and described later in this Example. These OsRAP1B interactors include prey clones of
5 OsMADS45. A BLAST analysis comparing the nucleotide sequence of OsRAP1B against TMRI's GeneChip[®] Rice Genome Array sequence database identified probeset OS003005.1_I_at ($2e^{-82}$ expectation value) as the closest match. Gene expression analysis indicated that this gene is expressed in roots and leaves and more highly expressed in flowers, panicles, and seeds. The gene is not specifically induced by a broad
10 range of plant stresses, herbicides and applied hormones.

Two-hybrid system using OsRAP1B as bait

Bait constructs containing the *O. sativa* AP1-like MADS box protein RAP1B (OsRAP1B) were constructed to search for interacting proteins. This protein is described
15 in earlier in this Example as an interactor for OsMADS45. Several bait fragments were used in the search encompassing amino acids 1-150, 125-235, 1-247, 100-247, 65-200, and 30-180 of OsRAP1B (see Table 9).

A bait encoding amino acids 1-150 of OsRAP1B was found to interact with a fragment of the transcription factor Os008339. This protein is described earlier in this
20 Example as an interactor for the bait protein OsMADS45. The Os008339-OsRAP1B interaction has not been previously reported.

A bait encoding amino acids 125-235 of OsRAP1B was also found to interact with rice MADS box-like protein OsBAA81880. OsBAA81880 (GenBank Accession
25 No. AB003322) is a 228-amino acid protein with a MADS box domain between amino acids 1 and 60 ($4.59e^{-36}$), as determined by amino acid sequence analysis. The analysis also detected two coiled-coil signatures (amino acids 83 to 113 and amino acids 140 to 174). These putative coiled coil regions overlap with a K-box domain (amino acids 73 to 173, $3.80e^{-32}$). The OsBAA81880 protein is not described in the literature; however, the
30 presence of the MADS box and K-box strongly suggests that it is a transcription factor of

the MADS box family, and its interaction with OsRAP1B is likely involved in transcriptional regulation of genes associated with plant development.

OsBAA81880 was also found to interact with OsMADS6 (see Table 10). A BLAST analysis comparing the nucleotide sequence of OsBAA81880 against TMRI's
5 GeneChip® Rice Genome Array sequence database identified probeset OS011977_i_at and OS011794_i_at (e^{-25} and e^{-12} expectation values, respectively) as the closest matches. The expectation values are too low for these probesets to be reliable indicators of the gene expression of OsBAA81880.

10 Baits encoding amino acids 1-247 of OsRAP1B and amino acids 100-247 of OsRAP1B were also found to interact with rice MADS-box protein FDRMADS6. This protein is described in earlier in this Example as an interactor for the bait protein OsMADS45. The OsFDRMADS6-OsRAP1B interaction has not been previously reported.

15 Baits encoding amino acids 1-247 of OsRAP1B and amino acids 100-247 of OsRAP1B was also found to interact with rice MADS box protein OsFDRMADS8. This protein is described earlier in this Example as an interactor for the OsMADS45 bait protein. The OsFDRMADS8-OsRAP1B interaction represents a heterodimer that has not been previously reported.

20 Baits encoding amino acids 1-247 of OsRAP1B, amino acids 100-247 of OsRAP1B, amino acids 65-200 of OsRAP1B, and amino acids 125-235 of OsRAP1B was also found to interact with MADS box protein OsMADS1. This protein is described herein as an interactor for the OsMADS45 bait protein. The OsMADS1-OsRAP1B interaction has not been previously reported.

25 Baits encoding amino acids 30-80 of OsRAP1B, amino acids 1-247 of OsRAP1B, amino acids 125-235 of OsRAP1B were also found to interact with rice MADS box protein OsMADS5. This protein is described herein as an interactor for the OsMADS45 bait protein. The OsMADS5-OsRAP1B interaction has not been previously reported.

A bait encoding amino acids 1-247 of OsRAP1B was also found to interact with rice MADS box protein OsMADS6. This protein is described earlier in this Example as an interactor for the OsMADS45 bait protein. The OsMADS6-OsRAP1B interaction has not been previously reported.

5

A bait encoding amino acids 1-247 of OsRAP1B was also found to interact with rice MADS box protein OsMADS7. OsMADS7 (GenBank Accession No. U78891) is a 259-amino acid protein with a MADS box domain between amino acids 11 and 71 ($3.22e^{-40}$), as predicted by analysis of the amino acid sequence. The analysis also predicted two coiled-coil signatures (amino acids 93 to 126 and 162 to 186). These coiled coils do not overlap with the MADS box domain. OsMADS7, as well as OsMADS8, is structurally related to the AGL2 gene family based on sequence homology and is a flower-specific MADS box gene (Kang *et al.*, *Mol. Cells* 7: 559-66, 1997). Both genes are expressed from the young flower stage through the late stage of flower development, with transcripts detected primarily in carpels and also weakly in anthers (Kang *et al.*, *surpa*). In support of an important role for OsMADS7 in flower development, specifically, in controlling flowering time, transgenic tobacco plants engineered to express the OsMADS7 gene were observed to exhibit early flowering and dwarfism (Kang *et al.*, *surpa*). The OsMADS7-OsRAP1B interaction has not been previously reported.

20

OsMADS7 was also found to interact with OsMADS6 (see Table 10). A BLAST analysis comparing the nucleotide sequence of OsMADS8 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS014912_f_at (e^{-61} expectation value) as the closest match. Gene expression analysis indicated that this gene is expressed early in seed development and is not specifically induced by a broad range of plant stresses, herbicides and applied hormones.

25

Baits encoding amino acids 1-247, 30-180, 100-247, and 125-235 of OsRAP1B were found to interact with rice MADS box protein OsMADS8. OsMADS8 (GenBank Accession No. U78892) is a 248-amino acid protein that includes a MADS box domain (amino acids 1 to 61, $3e^{-40}$), as determined by amino acid sequence analysis. Thus,

30

OsMADS8 is a member of the MADS box protein family. The amino acid sequence analysis also predicted a coiled coil C-terminal to the MADS box domain (amino acids 87 to 117). This coiled coil is likely part of a K-box predicted between amino acids 73 and 176 ($8.9e^{-44}$ prediction value). OsMADS8, as well as OsMADS7, is structurally
5 related to the AGL2 gene family, as determined by sequence homology, and is a flower-specific MADS box gene (Kang *et al.*, *Mol Cells* 7(4): 559-66, 1997). Both genes are expressed from the young flower stage through the late stage of flower development, with transcripts detectable primarily in carpels and also weakly in anthers (Kang *et al.*,
10 *supra*). In support of an important role for OsMADS7 and OsMADS8 in flower development, specifically, in controlling flowering time, is the observation that transgenic tobacco plants engineered to express these genes exhibit early flowering and dwarfism (Kang *et al.*, *supra*). The OsMADS8-OsRAP1B interaction represents a heterodimer that has not been previously reported.

OsMADS8 was also found to interact with the bait proteins OsMADS6 (see Table
15 10) and OsMADS3 (see Table 12). A BLAST analysis comparing the nucleotide sequence of OsMADS8 against TMRI's GeneChip[®] Rice Genome Array sequence database identified probeset OS015209_at (e^{-83} expectation value) as the closest match. Analysis of temporal and spatial patterns of gene expression indicated that this gene is expressed early in seed development. Analysis of gene expression in response to various
20 inducers indicated that it is not specifically induced by a broad range of plant stresses, herbicides and applied hormones.

A bait encoding amino acids 1-247 of OsRAP1B was found to interact with rice MADS box protein OsMADS17. OsMADS17 (GenBank Accession No. AF109153) is a
25 249-amino acid protein that includes a MADS box domain (amino acids 1 to 61), as determined by amino acid sequence analysis ($4.31e^{-41}$ prediction value). Thus, OsMADS17 is a member of the MADS box protein family. The amino acid sequence analysis also predicted a coiled coil located C-terminal to the MADS box domain (amino acids 122 to 178). This predicted coiled coil is likely part of a K-box predicted between
30 amino acids 72 and 174 ($5.2e^{-44}$). The OsMADS17 gene is homologous to ZAG3, the maize homolog of *Arabidopsis* AG, and belongs to the AGL6 subfamily in the

AP1/AGL9 family of MADS box genes (Moon *et al.*, *supra*). The OsMADS17-OsRAP1B interaction represents a heterodimer that has not been previously reported. The prey clone of OsMADS17 retrieved in the screen includes the predicted coiled coil and most of the K-box in OsMADS17.

5 OsMADS17 was also found to interact with the bait protein OsMADS5 (see Table 13). An interaction of OsMADS17 with OsMADS6 has also been reported (Moon *et al.*, *supra*). A BLAST analysis comparing the nucleotide sequence of OsMADS8 against TMRI's GeneChip[®] Rice Genome Array sequence database identified probeset OS000571_f_at (e^{-60} expectation value) as the closest match. Analysis of gene
10 expression indicated that this gene is not specifically induced by a broad range of plant stresses, herbicides and applied hormones.

Baits encoding amino acids 1-247, 30-180, and 125-235 of OsRAP1B were also found to interact with the rice MADS box protein OsMADS45, as has described earlier in
15 this Example. This interaction confirms the interaction between the two proteins used in the reverse bait/prey roles in the yeast two-hybrid system (see Table 1).

A bait encoding amino acids 1-247 of OsRAP1B was also found to interact with novel protein OsPN22834, a protein sharing similarity with Oshox6. OsPN22834 is a
20 278-amino acid protein that includes a homeobox domain between amino acids 70 and 131, a transposase 8 domain between amino acids 1 and 93, and a bZIP transcription factor domain between amino acids 129 and 167. Hox genes are well defined as modulators of development and pattern formation in a variety of species and organ systems (Fromental-Ramain *et al.*, *Development* 122: 461-472, 1996; Godwin *et al.*,
25 *Proc. Natl. Acad. Sci. USA* 95: 13042-13047, 1998). These genes code for transcription factors that modulate expression of developmentally regulated genes. While most of the published studies pertaining to Hox proteins utilize mouse models, Hox gene products have also been shown to regulate development in plants (Holk *et al.*, *Plant Mol. Biol.* 31: 1153-1161, 1996). The OsRAP1B- OsPN22834 interaction represents a previously
30 unreported heterodimer of a MADS box protein with a hox gene product.

Two-hybrid system using OsMADS6 as bait

O. sativa MADS box protein MADS6 was also used as a bait protein to identify interactors. This protein is described earlier in this Example as an interactor for the bait protein OsMADS45. The bait fragment used in this search encodes amino acids 50 to
5 200, a sequence that includes the predicted coiled coil and the K-box of OsMADS6.

OsMADS6 was found to interact with *O. sativa* OS008339 MADS box transcription factor (Os008339). This protein is described earlier in this Example as an interactor for the bait protein OsMADS45. The Os008339-OsMADS6 interaction
10 represents a newly identified interaction that is likely involved in transcriptional regulation of genes associated with development in rice.

OsMADS6 was also found to interact with the *O. sativa* MADS box-like protein OsBAA81880. This protein is described earlier in this Example as an interactor for the
15 bait protein OsRAP1B. The OsBAA81880-OsRAP1B interaction represents a heterodimer that has not been previously reported.

OsMADS6 was also found to interact with *O. sativa* MADS-box protein OsFDRMADS8. This protein is earlier in this Example as an interactor for the bait
20 protein OsMADS45. The OsFDRMADS8- OsMADS6 interaction has not been previously reported.

OsMADS6 was also found to interact with *O. sativa* MADS box protein OsMADS1. This protein is described earlier in this Example as an interactor for the bait
25 protein OsMADS45. This interaction confirms a previous work by Moon *et al.* (*Plant Physiol.* 120(4): 1193-204, 1999) who identified the same interaction using a yeast two-hybrid system.

OsMADS6 was also found to interact with *O. sativa* MADS box protein
30 OsMADS5. This protein is described earlier in this Example as an interactor for the bait

protein OsMADS45. This interaction confirms a previous work by Moon *et al. (supra)* who identified the same interaction using a yeast two-hybrid system.

OsMADS6 was also found to interact with *O. sativa* MADS box protein
5 OsMADS7. This protein is described earlier in this Example as an interactor for the bait protein OsRAP1B. This interaction confirms a previous work by Moon *et al. (supra)* who identified the same interaction using a yeast two-hybrid system.

OsMADS6 was also found to interact with *O. sativa* MADS box protein
10 OsMADS8. This protein is described earlier in this Example as an interactor for the bait protein OsRAP1B. This interaction confirms a previous work by Moon *et al. (supra)* who identified the same interaction using a yeast two-hybrid system.

OsMADS6 was also found to interact with *O. sativa* MADS box protein
15 OsMADS15. This protein is described earlier in this Example as an interactor for OsMADS45. Its interaction with OsMADS6 confirms a previous work by Moon *et al. (supra)* who identified the same interaction using the yeast two-hybrid system.

OsMADS6 was also found to interact with *O. sativa* MADS box protein
20 OsMADS18. This protein is described earlier in this Example as an interactor for OsMADS45. Its interaction with OsMADS6 confirms a previous work by Moon *et al. (supra)* who identified MADS18, as well as MADS14, MADS15, and MADS17, as interactors for MADS6 using the yeast two-hybrid system.

25 OsMADS6 was also found to interact with *O. sativa* MADS box protein OsMADS45. This protein is described earlier in this Example as a bait. The OsMADS45- OsMADS6 interaction confirms the interaction observed using OsMADS45 as bait, and represents a newly identified interaction.

30 OsMADS6 was also found to interact with novel protein OsPN29949. OsPN29949 is a novel 241-amino acid protein that includes a MADS box domain (amino

acids 1-61). The presence of this domain suggests that this protein is a member of the MADS box protein family. The alignment analysis of the interacting clones (see Figures 3A and 3B) shows that OsPN29949 shares high sequence similarity with OsMADS18, a member of the SQUA subfamily of AP1-like MADS box proteins. OsPN29949 may thus be classified in this group of genes, which are known to be involved in specification of floral organ primordia in snapdragon (reviewed in Moon *et al.*, *supra*). The OsPN29949-OsMADS6 interaction represents a newly identified heterodimer that is likely involved in transcriptional regulation of genes associated with development in rice.

Two prey clones encoding amino acids 118-241 and 109-193 of OsPN29949 were retrieved in the screen. These sequences suggest that the domain responsible for the OsPN29949-OsMADS6 interaction resides between amino acids 118 and 193, which includes the K box (amino acids 95-169) (see alignment analysis in Figure 3A). There is no match for the OsPN29949 gene on TMRI's GeneChip® Rice Genome Array.

OsMADS6 was also found to interact with *O. sativa* AP-like MADS box protein OsRAP1B. This protein is described earlier in this Example as an interactor for the bait protein OsMADS45, and was also used as a bait whose interactions are also reported earlier in this Example. The OsRAP1B-OsMADS6 interaction represents a heterodimer that has not been previously reported.

OsMADS6 was also found to interact with *O. sativa* Prolamin (OsRP5). Prolamin (GenBank Accession Nos. AF156714, AAF73991) is a 156-amino acid protein with a cleavable signal peptide domain (amino acids 1-19), as determined by analysis of the amino acid sequence. Prolamins are seed storage proteins unique to the endosperm of cereals. Seed storage proteins consist of polypeptide chains that are synthesized during seed development and serve as the main source of amino acids for germination and seedling growth. Prolamins accumulate in protein bodies derived from the endoplasmic reticulum (ER). The presence of the cleavable signal peptide domain in OsRP5 is consistent with the structure of prolamins, which possess signal peptides that direct the newly translated polypeptides into the lumen of the ER and are then proteolytically removed. In the ER, prolamins form aggregates and subsequently pinch off to form

protein bodies surrounded by an ER-derived membrane (the molecular structure of seed storage proteins and the mechanisms for their delivery into the vacuoles in seeds are discussed in Biochemistry and Molecular Biology of Plants, Buchanan, Gruissem and Jones (eds.), John Wiley & Sons, New York, NY 2002). The OsRP5-OsMADS6
5 interaction represents a previously unreported heterodimer.

In addition to OsMADS6, the prolamin OsRP5 was found to interact with rice hypothetical protein Os006111-3329, which is similar to the zinc/DNA-binding ascorbate oxidase promoter binding protein (AOBP) from *Curcubita maxima* and which includes a
10 Dof domain zinc finger DNA-binding domain (amino acids 103 to 165, $1.9e^{-37}$). The presence of the Dof domain suggests that Os006111-3329 is a transcriptional regulator. The interaction of prolamin with this protein and with OsMADS6 may represent steps in the transcriptional regulation of genes controlling seed development.

A BLAST analysis comparing the nucleotide sequence of prolamin against TMRI's GeneChip® Rice Genome Array sequence database identified probeset
15 OS000235 _at (e^{-155} expectation value) as the closest match. Analysis of gene expression indicated that this gene is not specifically induced by a broad range of plant stresses, herbicides and applied hormones.

20 Two-hybrid system using OsFDRMADS8 as bait

Two-hybrid assays were also performed using the *O. sativa* MADS-box protein FDRMADS8 as bait. This protein is described earlier in this Example as an interactor for the bait protein OsMADS45. The bait clone used in the screen encodes amino acids 60 to 160 of OsFDRMADS8.

25 OsFDRMADS8 was found to interact with OsMADS45. This protein is described as a bait earlier in this Example. The OsFDRMADS8-OsMADS45 interaction confirms the interaction between the two proteins used in the reverse bait/prey roles in the yeast two-hybrid system.

30

Two-hybrid system using OsMADS3 as bait

Two-hybrid assays were also performed using *O. sativa* MADS box protein MADS3 as bait. This protein is described earlier in this Example as an interactor for the bait protein OsMADS45. The bait clone used in the screen encodes amino acids
5 70 to 170 of OsMADS3.

OsMADS3 was found to interact with MADS box protein OsMADS8. This protein is described earlier in this Example as an interactor for the bait protein OsRAP1B. The OsMADS8-OsMADS3 interaction has not been previously reported.
10

OsMADS3 was also found to interact with OsMADS45. This protein is described as a bait earlier in this Example. The OsMADS45-OsMADS3 interaction confirms the interaction between the two proteins used in the reverse bait/prey roles in the yeast two-hybrid system.
15

OsMADS3 was also found to interact with OsPN31165, a novel 301-amino acid protein similar to three proteins of unknown function from *A. thaliana* (the first hit being unknown protein, GenBank Accession No. NP_565966, 62% identity; $2e^{-087}$), as determined by BLAST analysis. While the function of OsPN31165 is unknown, its
20 association with OsMADS3 suggests a role for OsPN31165 in plant development, most likely flower development. The OsMADS3-OsPN31165 interaction represents a newly identified heterodimer.

Two-hybrid assay using OsMADS5 as bait

Two hybrid assays were also performed using OsMADS5 as bait. This protein is described earlier in this Example as an interactor for OsMADS45. The bait clone used in the screen encodes amino acids 50 to 160 of OsMADS5.
25

OsMADS5 was found to interact with OsFDRMADS6. This protein is described earlier in this Example as an interactor for OsMADS45. The OsFDRMADS6-OsMADS5 interaction represents a heterodimer that has not been previously reported.
30

OSMADS5 was found to interact with OSMADS13. This protein is described earlier in this Example as an interactor for OSMADS45. The OSMADS13-OSMADS5 interaction has not been previously reported.

5

OSMADS5 was also found to interact with OSMADS17. This protein is described earlier in this Example as an interactor for OsRAP1B. The OSMADS17-OSMADS5 interaction has not been previously reported.

10

OSMADS5 was also found to interact with hypothetical protein 000564-1102 (Os000546-1102). Os000564-1102 is a novel 262-amino acid protein similar to the 14-3-3-like homolog GF14-b protein from rice (GenBank Accession No. AAB07456.1; 98% identity; $1e^{-141}$), as determined by BLAST analysis. 14-3-3 proteins include two highly conserved signature patterns: the first is a peptide of 11 amino acids located in the N-terminal section; the second is a 20-amino acid region located in the C-terminal section. Amino acid sequence analysis of Os000564-1102 identified a 14-3-3 signature 1 beginning with amino acid 49 and a 14-3-3 signature 2 beginning with amino acid 221. The 14-3-3 family members interact with, and thereby regulate, proteins that are involved in a variety of signaling pathways including transcriptional regulation. It is likely that Os000564-1102 is a 14-3-3 protein that regulates nuclear events such as transcription by participating in protein-protein interactions. Given the involvement of OSMADS5 in flower development, the interaction between OSMADS5 and Os000564-1102 likely represents a newly identified heterodimer involved in control of transcriptional events associated with plant development, and that Os000564-1102 modulates the MADS box transcription factor function as a member of the 14-3-3 family.

25

OSMADS6 was also found to interact with rice hypothetical protein BAB56078. This protein is a direct submission to the public domain (GenBank Accession No. BAB56078) and is not described in the literature. However, its association with OSMADS5 suggests a role for OsBAB56078 in plant development and this association represents a heterodimer that has not been previously reported.

30

OsBAB56078 was also found to interact with the rice 14-3-3 protein homolog GF14-b (OsGF14-b), which is up-regulated by stress and the plant hormone abscisic acid (as determined by gene expression analysis) (see Example V), and with the transcription
5 factor NAC2 (OsORF01393-P14).

Two-hybrid assays using OsMADS15 as bait

Two-hybrid assays were also performed using OsMADS15 as bait. This protein is described earlier in this Example as an interactor for OsMADS45. The bait clone used
10 in the screen encodes amino acids 100 to 235 of OsMADS15.

OsMADS15 was found to interact with MADS box protein OsMADS1. This protein is described herein as an interactor for OsMADS45. The OsMADS1-OsMADS15 interaction confirms a previous work by Lim *et al.* (*Plant Mol. Biol.* (2000) 44:513-27),
15 who identified OsMADS15 as well as OsMADS14 as interactors for OsMADS1 using the yeast two-hybrid system and determined that, while the K domain is essential for the interaction between these proteins, a region preceded by the K domain augments this interaction.

20 OsMADS15 was also found to interact with OsMADS45. This protein is described herein as a bait protein. The OsMADS45-OsMADS15 interaction confirms the interaction between the two proteins used in the reverse bait/prey roles in the yeast two-hybrid system.

25 OsMADS15 was also found to interact with OsPN29971, a 108-amino acid protein determined by BLAST analysis to be similar to centromere protein-like from *A. thaliana* (GenBank Accession No. 191066.1; 31.1% identity; $9e^{-09}$). The centromere is a region of the chromosome associated with kinetochores, protein-rich structures that are the main sites of interaction between cytoskeletal structures and chromosomes during
30 mitosis and meiosis. Centromere proteins in animals have been implicated in chromosome segregation and cytokinesis events. OsPN29971 may represent a novel

centromere-kinetochore-associated protein in plants. Its association with the MADS box protein OsMADS15 represents a newly identified heterodimer that likely regulates transcriptional events related to cell division during plant development.

5 Summary

The interacting proteins isolated in the two-hybrid screen using OsMADS45, OsRAP1B, and OsMADS6 as baits form a network comprised mainly of MADS box transcription factors. This indicates that MADS box proteins efficiently interact with each other in yeast, as previously reported (Moon *et al.*, *supra*).

10

Among the interactors found are the previously identified MADS box proteins Os008339, OsFDRMADS6, OsFDRMADS8, OsMADS1, OsMADS3, OsMADS5, OsMADS6, OsMADS7, OsMADS8, OsMADS13, OsMADS14, OsMADS15, OsMADS17, OsMADS18, OsBAA81880, OsMADS45, OsRAP1B and OsMADS6, and
15 the novel protein OsPN29949 (which interacted with OsMADS6). Because MADS box proteins are known to mediate various plant developmental processes as heterodimers, and given the involvement of the bait proteins OsMADS45, OsRAP1B and OsMADS6 in the regulation of flower development, the interactions between the MADS box proteins identified in this Example likely represent a network of heterodimers that regulate
20 transcription of genes associated with plant development in rice. Some of these interactions represent previously unreported heterodimers, as indicated in the description of each interactor in Sections 1-7.

Five additional novel interactors were identified: OsPN23495 is a putative
25 transcriptional regulator that, by association with OsMADS45, is also likely involved in flower development. OsPN22834 is a putative hox gene product. Both MADS box proteins and Hox gene products are well known for their roles in developmental processes, MADS box proteins being linked to flower and fruit development and Hox proteins to embryonic development in plants (Holk *et al.*, *Plant Mol. Biol.* 31(6): 1153-
30 1161, 1996). The interaction between RAP1B and OsPN22834 may signify a previously unknown role for one or both of these proteins in the development of the rice plant.

Os000564-1102 is a putative 14-3-3 protein that presumably modulates the function of the MADS box transcription factor OsMADS5 with which it interacts. OsPN29971 is a protein whose similarity to a centromere-like protein from *Arabidopsis* (although with low prediction significance) suggests a role in cell division events. The interaction of
5 OsPN29971 with the MADS box protein OsMADS15 is likely involved in regulating transcription of genes during cell division events related to plant development. Finally, OsPN31165 is a protein of unknown function, which by virtue of its interaction with OsMADS3 is likely involved in regulation of plant developmental processes. The
10 association of these novel interactors with the MADS box bait proteins of this Example represent newly identified heterodimers.

Another newly characterized heterodimer reported in this Example is that between OsMADS6 and the seed storage protein prolamin (OsRP5). Expression of storage
15 proteins and timing of their appearance in developing seeds is regulated both transcriptionally and post-transcriptionally. Regulatory sequences have been identified that control their temporal and spatial expression and determine seed and tissue
specificity, and more than one regulatory region (promoter) in the storage protein genes is thought to be involved in such regulation by specific DNA-binding proteins
(Biochemistry and Molecular Biology of Plants, Buchanan, Gruissem and Jones (eds.),
20 John Wiley & Sons, New York, NY 2002). The prolamin OsRP5 was found to interact with OsMADS6 and with another transcriptional regulator (not included in this
Example). It is possible that these interactions represent steps in the transcriptional regulation of prolamin expression associated with seed development. Alternatively, the
MADS box protein may be sequestered through the interaction with prolamin to be stored
25 with storage proteins that will be used upon seed germination. In either case, this interaction signifies a previously unreported role for OsMADS6 in seed development, in addition to flower development.

It is likely that the coiled coil(s)/K-box identified in the MADS box proteins of
30 this Example facilitate the MADS box protein interactions. Our amino acid sequence alignment analysis of the regions encoded by the interacting clones indicates that all

clones share a highly conserved MADS domain, a less conserved K box, and the more variable I region (directly downstream of the MADS domain) and C-terminal domain, in accordance with the modular structure reported in the literature for MADS box proteins (Moon *et al.*, *Plant Physiol.* 120(4): 1193-1204, 1999; Lim *et al.*, *Plant Mol. Biol.* 44(4): 513-527, 2000). The alignments are shown in Figure 3A. This analysis also determined that all interacting fragments include at least the K box, suggesting that this domain is responsible for dimerization, as reported previously. Furthermore, from these alignments a phylogenetic tree was constructed (shown in Figure 3B) to illustrate the relationships among the interacting proteins. Based on previous reports (Moon *et al.*, *supra*), the tree indicates that OsMADS45, OsMADS7, OsMADS8, OsMADS1 and OsMADS5 are members of the AGL2 subfamily; OsMADS6 and OsMADS17 belong to the AGL6 subfamily; OsFDRMADS6, OsMADS14, RAP1B, OsMADS15, OsMADS18 and novel protein OsPN29949 belong to the SQUA subfamily, all these subfamilies comprised in the AP1/AGL9 family of MADS box genes. The remaining interactors—OsMADS13, OsMADS3, OsFDRMADS8, OsBAA81880, and Os008339—are classified as others.

MADS box genes isolated from several plant species are known to play important roles in plant development, especially flower development. Knowledge of genes that regulate developmental processes such as flower and fruit development and flowering time has important applications in agriculture, providing new approaches to control of flower and fruit yield. For example, a mutant MADS-box gene, the apple PI homolog (MdPI) of the Arabidopsis mutant PI (which causes apetaly) abolishes the normal expression of the MdPI gene, resulting in parthenocarpic fruit (fruit without seed) development in some apple varieties (Yao *et al.*, *Proc. Natl. Acad. Sci. USA* 98(3): 1306-1311, 2001). Parthenocarpic fruit develops without pollination or fertilization and has a higher commercial value than its seed-bearing counterpart. The identification of the MdPI sequence has led to the proposal of genetic engineering methods to produce parthenocarpic fruit cultivars.

As one of the major human staples, rice has been a target of genetic engineering for higher yields and resistance to diseases, pests, and environmental stresses of various

kinds. The proteins encoded by MADS genes regulate transcription of genes associated with developmental processes such as floral organ identity, flowering time, and fruit development. The interactions between rice MADS box transcription factors identified in this Example are relevant to agriculture. Modulation of these interactions may be exploited for the development of genetically engineered plants characterized by a modulated flower development. Because rice is a model for other cereals, knowledge of the genetic mechanisms controlling development in rice will lead to opportunities for enhanced food crops.

The timing of the transition from vegetative growth to flowering, for example, is one of the most important steps in plant development. This step determines the quality and quantity of most crop species by affecting the balance between vegetative and reproductive growth. Therefore, control of flowering time in genetically engineered cereal crops is important in agriculture. One genetic modification that would be economically desirable would be to accelerate the flowering time of a plant. Induction of flowering is often the limiting factor for growing crop plants. One of the most important factors controlling induction of flowering is day length, which varies seasonally as well as geographically. There is a need to develop methods for controlling and inducing flowering in plants, regardless of the locale or the environmental conditions, thereby allowing production of crops, at any given time. Since most crop products (*e.g.*, seeds, grains, fruits), are derived from flowers, such a method for controlling flowering would be economically invaluable. A gene that modulates flowering time in plants was identified and its use proposed for the production of genetically modified plants in which overexpression of this gene results in early flowering in *Arabidopsis*, while loss of function mutations in or antisense directed to the gene cause late flowering (see U.S. Patent Application No. US20010049831 A1). Isolated nucleic acids and methods related to the OsMADS1, OsMADS5, OsMADS6, OsMADS7, and OsMADS8 genes of *Oryza sativa* and the NtMADS3 gene of *Nicotiana tabacum* have also been provided whose expression in transgenic plants causes an altered phenotype, including phenotypes related to the timing of the transition between vegetative and reproductive growth (*e.g.*, diminished apical dominance, early flowering, a partially or completely altered daylength

requirement for flowering, greater synchronization of flowering, or a relaxed vernalization requirement) (see U.S. Patent Application No. US5990386 A1).

Modulation of the protein interactions identified in this Example for OsMADS1, OsMADS5, OsMADS6, OsMADS7, and OsMADS8, for example, could lead to control
5 of flower induction in cereal crops. Additionally, modulation of plant development could be achieved through the identification and application of compounds that can affect the activity of the proteins or the expression of the genes provided in this Example.

In another potential application, the plant-specific K-box domain present in
10 MADS box proteins could be exploited for the development of compounds that increase the quantity or quality of fruit production but do not affect humans or livestock. Additionally, because the K-box domain is the region of the MADS box proteins that confers protein-binding specificity, these domains, either as parts or whole, can be targets
15 for genetic modification aimed at manipulating traits conferred by specific MADS box protein-protein interactions.

Example IV

Plant development may also be affected by proteins containing homeobox domains. As reviewed by Gehring, W.J., such homeobox domain containing proteins are
20 DNA-binding transcriptional regulators, many of which are involved in developmental processes (Gehring, W.J., *Trends Biochem. Sci.* 17(8): 277-280, 1992). Such proteins have been identified in plants (see, e.g., Ruberti *et al.*, *EMBO J.* 10(7): 1787-1791, 1991; Vollbrecht *et al.*, *Nature* 350(6315): 241-243, 1991). Homeobox genes are characterized
25 by the presence within each gene of a well-conserved sequence, the homeobox, which encodes a 61-amino acid DNA-binding domain called the homeodomain. The homeodomain-containing proteins encoded by the homeobox genes are thus capable of binding to specific DNA sequences and act as transcription factors that control the expression of downstream genes to regulate development. In higher plants,
homeodomain proteins are mainly implicated in organogenesis or developmental
30 processes (see references below), and also in the pathogenesis-related defense response (Korfhage *et al.*, *Plant Cell* 6: 695-708, 1994). The target genes directly regulated by

homeodomain-containing proteins are however still largely unidentified (Mannervick, *Bioessays* 21: 267-270, 1999).

Plant homeobox genes (reviewed in Chan *et al.*, *Biochim. Biophys. Acta* 1442: 1-19, 1998) can be subdivided into different families (Hd-Zip, Glabra, Knotted, PHD finger, Bell, Zmbox-PHD) according to sequence conservation within the homeodomain and the presence of additional sequences. Homeobox genes of the plant-specific knotted-like homeobox (KNOX) class contain a conserved domain, the KNOX domain, upstream of the homeodomain. The plant KNOX genes belong to the TALE superclass of homeobox genes, which also comprises genes identified in animals and fungi (Burglin *et al.*, *Nucleic Acids Res.* 25: 4173-4180, 1997). KNOX genes have been identified in numerous plants, both monocots such as rice and maize, and dicots such as *Arabidopsis* and tomato; they are normally expressed in the meristem and are thought to be primarily involved in shoot and leaf development, particularly in the control of cell fate determination in the shoot meristem (Chan *et al.*, *Biochim Biophys Acta* 1442(1): 1-19, 1998). The first identified plant homeobox gene, the KNOTTED1 (*kn1*; Vollbrecht *et al.*, *Nature* 350: 241-243, 1991) isolated from maize, provided evidence that plant homeobox genes, similar to those of animals, play an important role in regulating developmental processes. Ectopic expression of the maize *kn1* gene (and related dicot genes) often leads to the organization of new meristems in dicot leaves but usually not in monocot leaves (Haraven *et al.*, *Cell* 84(5): 735-44, 1996; Sinha *et al.*, *Genes Dev.* 7(5): 787-795, 1993; Lincoln *et al.*, *Plant Cell* 6(12): 1859-1876, 1994; Muller *et al.*, *Nature* 374(6524): 727-730, 1995; Williams-Carrier *et al.*, *Development* 124(19): 3737-3745, 1997; Hake *et al.*, *Philos Trans. R Soc. Lond. B Biol. Sci.* 350(1331): 45-51, 1995). Loss-of-function mutations in the maize *kn1* gene result in defects in shoot meristem maintenance (Kerstetter *et al.*, *Development* 124(16): 3045-3054, 1997). *Kn1* belongs to the plant-specific KNOX class of homeobox genes. Other KNOX genes identified in maize include *rough sheath1* (*rs1*) and *liguleless3* (*Lg3*) (reviewed in Chan *et al.*, *supra*; Muehlbauer *et al.*, *Plant Physiol.* 119(2): 651-62, 1999), which are thought to be involved in lateral organ development and specifically, in retarding the acquisition of terminal regional identity.

On the basis of sequence homology and expression pattern, KNOX genes are grouped into two classes, I and II (Kerstetter *et al.*, *supra*; Chan *et al.*, *supra*). Class I genes are mainly expressed in vegetative and inflorescence meristems and are involved in the regulation of shoot apical meristem formation and function and in leaf and flower morphology. The less characterized class II KNOX genes are expressed in most plant organs and tissues and not in meristematic tissues, and they are thought to regulate later stages of development. Further, all class I genes analyzed give rise to similar and distinct phenotypic effects, such as perturbations in the development of leaves leading to morphological defects, when ectopically expressed in transgenic plants. For example, the maize mutant *rough sheath2* (*rs2*) displays ectopic expression of at least three KNOX genes and consequently conditions a range of shoot and leaf phenotypes, including aberrant vascular development, ligular displacements, and dwarfism (Schneeberger *et al.*, *Development* **125**(15): 2857-2865, 1998). These studies suggest that down-regulation of KNOX gene expression is essential for normal leaf initiation and development. By contrast, no developmental defects have been recorded in plants expressing a class II gene ectopically.

Protein-protein interactions may contribute to the functioning of KNOX proteins, as demonstrated by the ability of two rice KNOX class I proteins to form homo- and heterodimers (Postma-Haarsma *et al.*, *Plant Mol Biol* **48**(4): 423-41, 2002). Besides the homeodomain (HD), KNOX proteins contain the conserved ELK and KNOX domains, the latter containing a putative helical structure that suggests a function in protein-protein interaction (Postma-Haarsma *et al.*, *supra*). In light of the importance of homeobox genes in controlling plant development, the interaction studies presented here are aimed at characterizing the rice homeobox protein OsHOS59, a member of the class II KNOX genes, which is not described in the literature. The identification of genes encoding proteins that participate in homeobox regulation in rice may allow genetic manipulation of crops to effect agronomically desirable changes in plant growth or development.

This Example provides newly characterized rice proteins interacting with the rice homeobox protein HOS59 (OsHOS59). An automated, high-throughput yeast two-hybrid

assay technology was used (provided by Myriad Genetics Inc., Salt Lake City, UT) to search for protein interactions with the bait protein OsHOS59.

Results

- 5 OsHOS59 was found to interact with five proteins annotated in the public domain: a hypothetical protein found similar to GTPase activating protein (OsAAD27557); a putative myosin (OsAAG13633); a putative homeodomain protein (OsAAK00972); putative eukaryotic translation initiation factor 3 large subunit; and the rice probable Myb factor. Seven additional interactors for OsHOS59 are novel rice proteins: a heat shock-
10 like protein (Os000221-3976); a protein similar to the rubber tree latex-abundant protein (OsPN23251); a putative S-adenosyl-L-homocysteine hydrolase (OsPN23829), an enzyme with a role in the control of methylation; a putative PHD-finger protein (OsPN23830); a myosin (OsPN24092) similar to the myosin protein OsAAG13633 described above; and two proteins of unknown function (OsPN23388 and OsPN30858).
15 Additional interactors were identified for some of the prey proteins.

The interacting proteins of the Example are listed in Table 15, followed by detailed information on each protein and a discussion of the significance of the interactions. The nucleotide and amino acid sequences of the proteins of the Example are
20 provided in Figure 11.

Some of the proteins identified represent rice proteins previously uncharacterized. Based on their presumed biological function and on the ability of the prey proteins to specifically interact with the bait protein OsHOS59, the interacting proteins are
25 speculated to be associated with developmental processes in rice.

Table 15. Interacting Proteins Identified for HOS59 (Homeobox Protein HOS59, Fragment).

30 The Myriad names and the TMRI names of the clones of the proteins used as baits and found as preys are given. Nucleotide/protein sequence accession numbers for the proteins of the Example (or related proteins) are shown in parentheses under the protein name. The bait and prey coordinates (Coord) are the amino acids encoded by the bait fragment(s) used in the search and by the interacting prey clone(s), respectively. The source is the library from which each prey clone was retrieved.

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (source)
-----------------------	--------------------------------------	------------	---------------------

BAIT PROTEIN :			
OsHOS59 PN20559	<i>O. sativa</i> Homeobox Protein HOS59, Fragment (BAB55659.1)		
INTERACTORS:			
OsAAD27557* PN22896	<i>O. sativa</i> Hypothetical Protein, Similar to GTPase Activating Protein (AF111710; AAD27557)	1-100	7-142 (input trait)
OsAAG13633# PN25701	<i>O. sativa</i> Putative Myosin (AC078840; AAG13633)	1-100	799-951 (output trait)
OsAAK00972 PN23253	<i>O. sativa</i> Putative Homeodomain Protein OsAAK00972 (AC079736; AAK00972.1)	1-100	236-350 (output trait)
OsBAB07943 PN23832	<i>O. sativa</i> Putative Eukaryotic Translation Initiation Factor 3 Large Subunit (AP002487; BAB07943.1)	1-100	525-767 (output trait)
OsMYB PN20689	<i>O. sativa</i> Probable Myb Factor (T03830)	1-100	36-129 (output trait)
Os000221-3976& PN23169	Hypothetical Protein 000221-3976, Fragment, Similar to OsHP82 (P33126; e=0.0)	1-100	2x 123-238 (input trait)
OsPN23251	Novel Protein PN23251	1-206	112-291 (input trait)
OsPN23388	Novel Protein PN23388	1-100	229-331 (output trait)
OsPN23829@	Novel Protein PN23829 Putative S- Adenosyl-L-Homocysteine Hydrolase (P32112; e=0.0)	1-100	3x 2-226 (output trait)
		1-206	3x 1-247 (output trait)
OsPN23830 !	Novel Protein PN23830, Similar to <i>A. thaliana</i> Putative PHD-Finger Protein (NP_566742.1; 2e ⁻⁷³)	1-100	4-207 2x 1-169 (output trait)
OsPN24092	Novel Protein PN24092, Similar to <i>O. sativa</i> Putative Myosin	1-100	797-948 (output trait)
OsPN30858	Novel Protein PN30058	1-206	230-400 (output trait)

* Additional interactions identified for OsAAD27557 are shown in Table 16

Additional interactions identified for OsAAG13633 are shown in Table 17

& Additional interactions identified for Os000221-3976 are shown in Table 18

@ Additional interactions identified for OsPN23829 are shown in Table 19

! Additional interactions identified for OsPN23830 are shown in Table 20

5

Table 16

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (source)
PREY PROTEIN:			
OsAAD27557 PN22896	Hypothetical Protein Similar to GTPase Activating Protein (AF111710; AAD27557)		
BAIT PROTEIN:			
Os003181-3684 PN21036	Hypothetical Protein 003181-3684	58-140	1-149 (output trait)

5

Table 17

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (source)
PREY PROTEIN:			
OsAAG13633 PN25071	<i>O. sativa</i> Putative Myosin (AC078840; AAG13633)		
BAIT PROTEIN:			
Os005750-3115 PN20466	<i>O. sativa</i> bZIP Transcription Factor (AB051294; BAB72061.1)	50-150	2x 528-789 538-738 612-738 (output trait)

Table 18

Myria/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (source)
PREY PROTEIN:			
Os000221-3976	Hypothetical Protein 000221-3976, Fragment, Similar to OsHP82 (P33126; e=0.0)		
BAIT PROTEIN:			
OsCYCOS2 PN20257	<i>Oryza sativa</i> Cyclin 2 (X82036; CAA57556)	50-233	163-313, (input trait)

10

Table 19

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (source)
PREY PROTEIN:			
OsPN23829	Novel Protein PN23829 Putative S-Adenosyl-L-Homocysteine Hydrolase (P32112; e=0.0)		
BAIT PROTEIN:			
OsTFX1 PN19697	<i>O. sativa</i> Putative Transcription Factor X1 (AF101045; AAF21887)	400-629	-21-216 -4-226 -2-195 (output trait)
Os005792-3529	Hypothetical Protein 005792-3529	1-55	3-220

PN20080	Similar to <i>O. sativa</i> Receptor Kinase (AAK18840.1; 8e ⁰⁷)	(output trait)
---------	---	----------------

Table 20

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (source)
PREY PROTEIN:			
OsPN23830	Novel Protein PN23830, Similar to <i>A. thaliana</i> Putative PHD-Finger Protein (NP_566742.1; 2e ⁻⁷³)		
BAIT PROTEIN:			
Os018049-3655 PN20534	Hypothetical Protein 018049-3655, Fragment, <i>O. sativa</i> Putative Homeodomain Transcription Factor, 3'-Partial (AC092697; AAL58126.1)	1-148	89-250 (output trait)

Two-hybrid assay using OsHOS59 as bait

- 5 OsHOS59 is a 205-amino acid protein fragment with a homeobox domain profile (Gehring W.J., *Trends Biochem. Sci.* **17**: 277-280, 1992; Gehring and Hiromi, *Ann. Rev. Genet.* **20**: 147-173, 1986; Schofield, P.N., *Trends Neurosci.* **10**: 3-6, 1987), namely at amino acids 122 to 185, as determined by analysis of its amino acid sequence. Proteins within this group are DNA-binding transcriptional regulators that are involved in
- 10 developmental processes. A BLAST analysis of the amino acid sequence indicated OsHOS59 is the rice KNOX Family Class II Homeodomain Protein (GenBank Accession No. BAB55659.1). The analysis indicated that all proteins displaying close homology to OsHOS59 are also homeodomain proteins, particularly from plant species. This strongly suggests that OsHOS59, although not described in the literature, is a rice homeobox
- 15 protein that most likely functions as do other members of this protein family.

- There is not much evidence on the role of class II KNOX genes. However, based on studies with the class II gene *KNAT3* from *Arabidopsis*, which was found to be expressed in young leaves, buds and pedicels, at the junction between organs and in
- 20 maturing tissues, and whose expression is regulated by light, class II KNOX genes are suggested to be involved in later stages of plant development (discussed in Chan *et al.*, *Biochim Biophys Acta* **1442**(1): 1-19, 1998).

Two bait fragments, encoding amino acid 1-100 and 1-206, of OsHOS59 were used in the yeast two-hybrid screen.

A BLAST analysis comparing the nucleotide sequence of OsHOS59 against
5 TMRI's GeneChip[®] Rice Genome Array sequence database identified probeset OS011682_at and OS002989.1_i_at (e^{-100} and $7e^{-26}$ expectation values, respectively) as the closest matches. Analysis of gene expression in rice plants indicated that this gene is down-regulated by environmental cold, and by abscisic acid and jasmonic acid.

10 OsHOS59 was found to interact with OsAAD27557. OsAAD27557 is annotated as a rice Hypothetical Protein (GenBank Accession No. AAD27557). It is a 789-amino acid protein with a leucine-rich repeat between amino acids 214 and 241, as determined by analysis of its amino acid sequence ($1.28e^{-03}$ prediction value). Leucine-rich repeats are thought to be involved in protein-protein interactions (Kobe *et al.*, *Trends Biochem.*
15 *Sci.* **19**: 415-421, 1994). A BLAST analysis against the public database indicated that the amino acid sequence of OsAAD27557 is similar to those of Ran GTPase activating protein from the plant *Medicago sativa* subsp. *x varia* (Accession #AAF19528.1, 66.4% identity, $e=0.0$) and GTPase activating protein 2 from *A. thaliana* (GenBank Accession No. NP_197433, 62% identity, e^{-179}). In agreement with these results, a BLAST analysis
20 against Myriad's proprietary database indicated human Ran GTPase activating protein 1 (RANGAP1) as the most similar protein to OsAAD27557 (28% identity, $5e^{-24}$). GTPase activating proteins interact with GTPases such as Ras thereby enhancing the GTPase activity (Bischoff, *et al.*, *Proc. Natl. Acad. Sci. USA* **91**: 2587-2591, 1994). Hydrolysis of GTP to GDP is an important step in many intracellular signal transduction pathways that
25 control various cellular processes such as cell growth and development, apoptosis, lipid metabolism, cytoarchitecture, membrane trafficking, and transcriptional regulation (Aznar and Lacal, *Prog Nucleic Acid Res. Mol. Biol.* **67**:193-234, 2001). Ran GTPases are required for nucleo-cytoplasmic transport, regulation of cell cycle progression, mitotic spindle formation, and postmitotic nuclear assembly (reviewed by Sazer, *et al.*, *J.*
30 *Cell Sci.* **113**(Pt 7): 1111-1118, 2000 and Dasso, M., *Cell* **104**: 321-324, 2000). Plants Ran proteins are thought to be functionally equivalent to their mammalian and yeast

homologs and to be necessary for maintaining a coordinated cell cycle, for protein import into the nucleus and for the onset of mitosis (Ach and Gruissem, *Proc. Natl. Acad. Sci. USA* 91: 5863-5867, 1997; Merkle *et al.*, *Plant J.* 6: 555-565, 1994). Moreover, plant small GTP-binding proteins have been linked to disease resistance (Ono *et al.*, *Proc. Natl. Acad. Sci. USA* 98: 759-764, 2001). Thus, the prey protein OsAAD27557 is a rice GTPase activating protein that likely participates in signal transduction involving GTP hydrolysis during events related to cell division as part of either plant development and/or response to pathogen invasion.

10 OsAAD27557 also interacts with Hypothetical Protein 003181-3684 (Os003181-3684) (see Table 16). Os003181-3684 is a hypothetical protein of 176 amino acids that includes a predicted transmembrane domain (amino acids 43 to 59). A BLAST analysis of the amino acid sequence indicated no proteins highly similar to Os003181-3684 in either public or Myriad's proprietary databases. However, the predicted transmembrane domain suggests that this protein may be some type of cell surface receptor or receptor-
15 interacting protein that is important for signal transduction. The OsAAD27557-Os0031813684 interaction may represent a step in a signal transduction pathway involving GTP hydrolysis and transcriptional regulation in developmental processes.

20 OsHOS59 was also found to interact with *O. sativa* putative myosin (OsAAG13633). A BLAST analysis of the amino acid sequence of OsAAG13633 indicated that this prey protein is the rice putative myosin (GenBank Accession No. AAG13633, 100% identity, $e=0.0$). Myosins are discussed in Example I. Based on current knowledge of plant myosins, the prey protein OsAAG13633 may be a
25 cytoskeletal component that participates in events relating to cytoplasmic streaming or cell division during plant development.

OsAAG13633 also interacts with *O. sativa* bZIP Transcription Factor (Os005750-3115) (see Table 17). Os005750-3115 is a 333-amino acid protein with a predicted basic leucine zipper (bZIP) domain (amino acids 45 to 108, $1.54e^{-6}$) (see Hurst, H.C.,
30 *Protein Prof.* 2: 105-168, 1995); Ellenberger, T., *Curr. Opin. Struct. Biol.* 4: 12-21, 1994). This domain includes a basic DNA-binding region and a leucine zipper used to

initiate protein-protein interactions, and it is often found in transcription factors. A BLAST analysis of the amino acid sequence of Os005750-3115 indicated that this protein is the rice bZIP Transcription Factor (GenBank Accession No. BAB72061.1, 99.3% identity, $e=0.0$).

5

OsHOS59 was also found to interact with OsAAK00972, a 642-amino acid protein that includes a homeobox domain profile (amino acids 379 to 442 by Prosite, amino acids 406 to 441 by Pfam), as determined by analysis of its amino acid sequence. The analysis also identified a POX domain (a domain associated with HOX domains) between amino acids 188 and 333 ($1.36e^{-56}$). The retrieved prey clone encodes amino acids 236 to 350 of OsAAK00972, a region that includes the POX domain of OsAAK00972. Hox genes are clustered sets of homeobox-containing genes that play a central role in animal development (Mann and Affolter, *Curr. Opin. Genet. Dev.* 8(4): 423-429, 1998). A BLAST analysis of the amino acid sequence of OsAAK00972 indicated that it is the rice Putative Homeodomain Protein (GenBank Accession No. AAK00972.1, 100% identity, $e=0.0$). OsAAK00972 is thus a member of the homeobox protein family.

OsHOS59 was also found to interact with OsBAB07943, a protein of 984 amino acids with a predicted transmembrane domain (amino acids 316 to 332). Analysis of its sequence also identified a PINT (Proteasome, Int-6, Nip-1 and TRIP-15) motif (amino acids 441 to 532, $3.91e^{-07}$), which is present in the C-terminal region of several regulatory components of the 26S proteasome and other proteins. The function of this motif is not known. The analysis also predicted three coiled coils (amino acids 91 to 123, 552 to 700, and 794 to 963). The prey clone retrieved encodes amino acids 525 to 767 of OsBAB07943, a region that includes one of the predicted coiled coils within OsBAB07943. The presence of the PINT motif is in agreement with the results of BLAST analysis, which indicated that OsBAB07943 is the rice putative eukaryotic translation initiation factor 3 (eIF3) large subunit (GenBank Accession No. BAB07943.1, 100% identity, $e=0.0$), eIF3e being homologous to the product of Int-6 (eIF3e) (Shalev *et al.*, *J. Biol. Chem.* 276: 34948-34957, 2001). The analysis also indicated that

OsBAB07943 is similar to eukaryotic translation initiation complexes of other species including *Zea mays* (gi 5106764, 69% identity, e=0.0) and *Nicotiana tabacum* (gi 6685538, 66% identity, e=0.0). Therefore, it is likely that OsBAB07943 truly is a rice translation initiation factor subunit.

5 The mammalian eukaryotic initiation factor 3 (eIF3) is composed of at least eight subunits, the largest of which has a relative molecular mass of 180 kDa. A comparison of the sequences of the corresponding eIF3 large subunits from several species led to the conclusion that eIF3 large subunit is highly conserved across the animal, plant, and fungal kingdoms (Johnson *et al.*, *J. Biol. Chem.* **272**: 7106-7113, 1997). In *Z. mays*,
10 eukaryotic translation initiation factor 3 large subunit is expressed in the region of the root meristem surrounding the central stele and in the young root, the male inflorescence, and the developing cob and seed (Sabelli *et al.*, *Mol. Gen. Genet.* **261**: 820-830, 1999). Eukaryotic initiation factor complexes initiate translation of mRNA (reviewed by Hannig *et al.*, *Bioessays* **17**: 915-919, 1995), in part by using their helicase activity to unwind the
15 mRNA strand secondary structure in the 5'-untranslated region of mRNA, which facilitates binding of the mRNA to the 40 S ribosomal subunit (Rogers *et al.*, *J. Biol. Chem.* **276**: 30914-30922, 2001). In addition, eIF3 in humans is in some circumstances regulated by protein-protein interaction (Guo *et al.*, *EMBO J.* **19**: 6891-6899, 2000).

20 OsHOS59 was also found to interact with *O. sativa* Myb factor (OsMYB). A BLAST analysis of the amino acid sequence of OsMYB indicated that this prey protein is the rice Probable Myb Factor (GenBank Accession No. T03830, 100% identity, e⁻¹⁶⁸). OsMYB is a protein of 279 amino acids that includes an ATP/GTP-binding site motif A (P-loop, amino acids 45 to 52; see, *e.g.*, Saraste *et al.*, *Trends Biochem. Sci.* **15**: 430-434, 1990); Koonin, E.V. *J. Mol. Biol.* **229**: 1165-1174, 1993) and two Myb DNA-binding
25 domain repeats (amino acids 17 to 25 for signature 1, and amino acids 89 to 112 for signature 2; see, *e.g.*, Grotewold *et al.*, *Proc. Natl. Acad. Sci. USA* **88**: 4587-4591, 1991; Oppenheimer *et al.*, *Cell* **67**: 483-493, 1991). The prey clone retrieved encodes amino acids 36 to 129 of OsMYB, a region that includes the P-loop and the Myb DNA-binding
30 domain signature 2. Myb proteins are nuclear DNA-binding proteins that recognize the sequence pyAAC(G/T)G (Biedenkapp, *et al.*, *Nature* **335**: 835-837, 1988). The presence

of two Myb DNA-binding signatures suggests that OsMYB is a member of the two-repeat family of Myb proteins. The number of these repeats determines how the protein binds DNA and, consequently, its function (reviewed by Jin and Martin, *Plant Mol. Biol.* 4: 577-585, 1999).

5

OsHOS59 was also found to interact with Os000221-3976, a 480-amino acid protein fragment that includes an Hsp90 domain (amino acids 6 to 480), as determined by analysis of its amino acid sequence ($e=0.0$). A BLAST analysis against the public and Myriad's proprietary databases showed that Os000221-3976 shares amino acid sequence
10 similarity with many heat shock proteins, the top hit being the rice heat shock protein 82 (Van Breusegem *et al.*, *Planta* 193(1): 57-66, 1994; GenBank Accession No. P33126, 96.4% identity, $e=0.0$). Therefore, Os000221-3976 is either a splice variant of heat shock protein 82 or a separate but very similar protein. A comparison of the nucleotide sequences suggests the latter is more likely. The rice HSP82 mRNA is induced
15 specifically upon heat stress (Van Breusegem *et al.*, *supra*).

While heat shock proteins (HSPs) have been ascribed a main role in the plant stress response, some of these proteins are designated as HSPs solely based on sequence homology and their functions in plants have not been demonstrated *in vitro*. Indeed, some HSPs are expressed throughout development. HSPs function as molecular
20 chaperones that promote proper protein folding and may have roles not related to the stress response. HSP70 proteins, for instance, are essential for normal cell function. They are ATP-dependent molecular chaperones that may interact with many different proteins, given their role in protein folding, unfolding, assembly, and disassembly. These topics are discussed in Biochemistry and Molecular Biology of Plants, Buchanan, Gruissem and
25 Jones (eds.), John Wiley & Sons, New York, NY 2002, pp.1197-1202. The heat shock protein HSP70 in sea urchin cells has been proposed to have a chaperone role in tubulin folding when localized on centrosomes, and in the assembling and disassembling of the mitotic apparatus when localized on the fibres of spindles and asters (Agueli *et al.*, *Biochem J.* 360: 413-419, 2001).

30 The heat shock protein Os000221-3976 also interacts with rice Cyclin 2 (OsCYCOS2) (see Table 18). The 419-amino acid protein OsCYCOS2 (GenBank

Accession No. CAA57556) is a G2/M type cyclin that contains two cyclin domains spanning amino acids 200 to 284 ($2.7e^{-26}$) and amino acids 297 to 379 ($1.29e^{-22}$). Type G2/M cyclins regulate the cell cycle progression from G2 to mitosis during plant development. Cyclins are regulatory proteins that activate cyclin-dependent protein
5 kinases (CDKs), which are essential for cell cycle progression in eukaryotes. The binding of cyclins to specific proteins is thought to provide potential substrates to CDKs. Cyclins are thus important regulators that couple control of proliferation to the many environmental and developmental cues that affect plant growth. (The role of cyclin-CDK complexes in regulation of the plant cell cycle is reviewed in John et al., *Protoplasma*
10 216:119-142, 2001 and Potuschak and Doerner, *Curr. Opin. Plant Biol.* 4: 501-506, 2001. Interactions identified for OsCYCOS2 are discussed in Example II above.)

OsHOS59 was also found to interact with OsPN23251, a novel 420-amino acid protein with a possible cleavage site between amino acids 19 and 20, although no N-
15 terminal signal peptide is evident. A BLAST analysis of the OsPN23251 amino acid sequence determined that it is similar to latex-abundant protein from the rubber tree *Hevea brasiliensis* (GenBank Accession No. AAD13216.1, 62% identity, e^{-141}). Many proteins isolated from latex are defense-related allergens (Kostyal, et al., *Clin. Exp. Immunol.* 112: 355-362, 1998)). A BLAST analysis comparing the nucleotide sequence
20 of OsPN23251 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset Os004430.1_at ($e=0.0$ expectation value) as the closest match. Analysis of gene expression indicated that this gene is specifically expressed in root.

OsHOS59 was also found to interact with novel protein OsPN23388. OsPN23388
25 is a 509-amino acid protein with a predicted BRCA1 C-terminus (BRCT) domain (amino acids 1 to 42, $5.2e^{-05}$), which is known to facilitate protein-protein interactions. This domain was originally identified in the breast/ovarian cancer suppression protein, BRCA1, and is found in a large number of proteins involved in DNA repair, recombination, and cell cycle control (Zhang et al., *EMBO J.* 17: 6404-6411, 1998).
30 These include p53-binding protein (53BP1) and two uncharacterized hypothetical proteins (KIAA0170 and SPAC19G10.7) (Callebaut and Mornon, *FEBS Lett.* 400: 25-30,

5

10

25

homeotic change in transgenic tobacco expressing antisense RNA of the S-adenosyl-L-homocysteine hydrolase gene (Tanaka *et al.*, *Plant Mol. Biol.* **35**: 981-986, 1997). In addition, a role for S-adenosyl-L-homocysteine hydrolase in the plant pathogen-induced defense response has been suggested based on the observation that elicitor treatment induces both S-adenosyl-L-homocysteine hydrolase mRNA expression and activity in parsley cultured cells and in intact leaves (Kawalleck *et al.*, *supra*). In a contrasting role, S-adenosyl-L-homocysteine hydrolase activity may be involved in mechanisms leading to viral infection, as the effectiveness of antiviral compounds correlates with their ability to inhibit its activity (Robins *et al.*, *J. Med. Chem.* **41**: 3857-3864, 1998; Liu *et al.*, *Antiviral Res.* **19**: 247-265, 1992; Wolf and Borchardt, *J. Med. Chem.* **34**: 1521-1530, 1991; Kitade *et al.*, *Nucleic Acids Symp. Ser.* **42**: 25-26, 1999).

15 A BLAST analysis comparing the nucleotide sequence of OsPN23829 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset Os001768.1_at (e=0.0) expectation value) as the closest match. Analysis of gene expression indicated that this gene is induced by jasmonic acid and by *Magnaporthe grisea*, the fungal pathogen that causes rice blast disease.

OsPN23829 also interacts with rice putative transcription factor X1 (OsTFX1) (GenBank Accession No. AAF21887.1), and with hypothetical protein 005792-3529 (Os005792-3529) (see Table 19). OsTFX1 is an uncharacterized transcription factor. It may form a complex with both OsPN23829 and OsHOS59 to regulate transcriptional events related to cell cycle/development. Os005792-3529 is a hypothetical protein of 54 amino acids in which no well-characterized protein domain was identified. The isolated cDNA sequence starts with the putative ATG initiation codon, leaving the reading frame potentially open in the 5' direction, suggesting that the real protein might be larger than 54 residues. BLAST analysis of the available amino acid sequence indicated that Os005792-3529 is similar to a putative receptor kinase from rice (Accession #AAK18840.1, 72% identity, $8e^{-07}$). (Note, however, that the domain of similarity with the putative receptor kinase AAK18840.1 is only 36-residue long.)

30 OsHOS59 was also found to interact with novel protein PN23830, which is similar to the putative *Arabidopsis* PHD-Finger protein OsPN23830. OsPN23830 is a

protein of 253 amino acids. An analysis of its amino acid sequence identified a PHD domain (plant homeo domain, Pascual *et al.*, *J. Mol. Biol.* 304: 723-729, 2000; Aasland *et al.*, *Trends Biochem. Sci.* 20: 56-59, 1995) (amino acids 199 to 246, e^{-10}). The presence of the PHD finger domain is in agreement with BLAST analysis which indicated
5 similarity of OsPN23830 to *Arabidopsis* putative PHD-finger protein (GenBank Accession No. NP_566742.1, 53.8% identity, $2e^{-73}$). The PHD finger is a Cys⁴-His-Cys³ zinc finger found primarily in a wide variety of chromatin-associated proteins, including HAT3.1, a plant homeobox gene (Aasland *et al.*, *supra*). Although the exact function of the PHD finger is not known, it is thought to facilitate protein-protein interactions
10 (O'Connell *et al.*, *J. Biol. Chem.* 276: 43065-43073, 2001). The association OsPN23830 with OsHOS59 suggests a role for OsPN23830 in transcriptional regulation during development.

OsPN23830 also interacts with another homeodomain protein, Hypothetical Protein 018049-3655 (Os018049-3655) (See Table 20). A BLAST analysis of the amino
15 acid sequence of Os018049-3655 determined that this protein is the rice Putative Homeodomain Transcription Factor, 3'-Partial) (GenBank Accession No. AAL58126.1, 100% identity, $5e^{-134}$).

OsHOS59 was also found to interact with novel protein PN24092. A BLAST
20 analysis of the amino acid sequence of OsPN24092 determined that this protein is similar to the same rice putative myosin (GenBank Accession No. AAG13633, 84.7% identity, $e=0.0$) found to interact with OsHOS59 (see *O. sativa* Putative Myosin (OsAAG13633)).

OsHOS59 was also found to interact with novel protein PN30058. A BLAST
25 analysis of the amino acid sequence of OsPN30858 determined that this protein is similar to Expressed Protein from *A. thaliana* (GenBank Accession No. NP_566372.1, 63.2% identity, $e=0.0$), a protein of unknown function.

30

Summary

The KNOX homeodomain protein OsHOS59 interacts with other DNA-binding proteins thought to be involved in transcriptional regulation, including a putative homeodomain protein (OsAAK00972) and a Myb protein (OsMYB). These interactions are consistent with published evidence that KNOX proteins function as homo- and heterodimers. Indeed, the specificity of KNOX proteins may be further enhanced by interactions with other transcription factors Mann and Affolter, *Curr. Opin. Genet. Dev.* 8: 423-429, 1998; Postma-Haarsma *et al.*, *Plant Mol. Biol.* 48: 423-441, 2002). Based on the presumed role of OsHOS59 in plant development, we speculate that the OsHOS59-OsAAK00972 and OsHOS59-OsMYB interactions represent protein complexes that regulate transcription of genes involved in developmental processes and, in the case of OsMYB regulation, which include a specific sequence in their promoters. This hypothesis is supported by the observation that both HOX and Myb transcription factors cooperatively function to regulate myeloid cell differentiation in mammals (Nagamara-Inoue *et al.*, *Int. Rev. Immunol.* 20: 83-105, 2001, and reviewed by Lenny *et al.*, *Mol. Biol. Rep.* 24: 157-168, 1997).

OsHOS59 was also found to interact with a putative Ran GTPase activating protein (OsAAD27557). Given the function of Ran GTPases in nucleo-cytoplasmic transport, regulation of cell cycle progression, mitotic spindle formation, and postmitotic nuclear assembly Sazer and Dasso, *J. Cell Sci.* 113(Pt 7): 1111-1118, 2000 and Dasso, *M. Cell* 104: 321-324, 2000), the OsHOS59-OsAAD27557 interaction is speculated to represent a step in a signal transduction pathway that involves GTP hydrolysis during events related to cell cycle progression or cell division as part either plant development and/or response to pathogen invasion.

Two of the interactors identified in the yeast two-hybrid screen, OsAAG13633 and the novel protein OsPN24092, are putative myosins highly similar to each other (84.7% identity). Note that OsAAG13633 also interacts with another transcription factor (Os005750-3115). Molecular motors, including kinesins, myosins and dyneins, have been well characterized in non-plant organisms and implicated in a variety of cellular functions such as vesicle and organelle transport, cytoskeleton dynamics, morphogenesis,

polarized growth, cell movements, spindle formation, chromosome movement, nuclear fusion, and signal transduction. In contrast, the roles of the many kinesins and myosins identified in plants are largely unknown (reviewed in Reddy, A.S. *Int. Rev. Cytol.* 204: 97-178, 2001). A few studies suggest that myosins in higher plants are involved in the movement of organelles and vesicles during cytoplasmic streaming and in pollen tube growth, and in maturation of the cell plate at cytokinesis (reviewed in Yokota *et al.*, *Plant Physiol.* 121:525-534, 1999; Reichelt *et al.*, *Plant J.* 19: 555-567, 1999). The rice myosins identified in this Example are likely involved in dynamic cytoskeletal events, such as cytoplasmic streaming, intracellular cargo movement or cell division, associated with development processes. Their interactions with the transcription factors OsHOS59 and Os005750-3115 may represent steps in transcriptional regulation of such events.

Another interactor, Os000221-3976, is a putative heat shock protein similar to rice HSP82. Heat shock proteins (HSPs) act as molecular chaperones and, while these molecules in plants have been mainly linked to the stress response, some are not related to stress and their functions remain to be defined (Biochemistry and Molecular Biology of Plants, Buchanan, Gruissem and Jones (eds.), John Wiley & Sons, New York, NY 2002, p. 1198). Indeed, some HSPs are expressed throughout development. In the context of all the interactions identified for OsHOS59, it is possible that Os000221-3976 acts as a molecular glue to hold together interacting proteins or to promote proper protein folding in events related to plant development which may or may not be associated with stress. An alternative role for this prey protein may be deduced by functional homology with animal heat shock proteins whose chaperone roles in tubulin folding or mitotic structures assembly/disassembly depends on their localization on centrosomes or spindle fibers, respectively (Agueli *et al.*, *Biochem J.* 360: 413-419, 2001). The heat shock protein Os000221-3976 may thus act as a chaperone in events related to tubulin folding or mitotic structure assembly/disassembly. These are functions associated with the phase of the cell cycle controlled by OsCYCOS2, a type G2/M cyclin that regulates the cell cycle progression from G2 to mitosis during plant development. The interaction identified in this Example between the heat shock protein Os000221-3976 and OsCYCOS2 substantiates this hypothesis and further supports the involvement of this

novel rice heat shock protein in developmental processes. Discovery of the subcellular localization of Os000221-3976 may clarify its function.

Another protein interacting with OsHOS59 with a role in regulation of development is a putative S-adenosyl-L-homocysteine hydrolase (OsPN23829), an enzyme involved in control of methylation reactions. Transmethylation reactions are important components of the biosynthetic machinery in most plant cells. S-adenosyl-L-homocysteine hydrolase participates in the activated methyl cycle which yields methionine, whose fate is important for protein synthesis or DNA modification. In plants, the regulation of intracellular methylation reactions mediated by S-adenosyl-L-homocysteine hydrolase has been linked to morphogenesis through *in planta* studies. Dereglulation of methylation results in morphological changes including a floral homeotic change in transgenic tobacco expressing antisense RNA of the S-adenosyl-L-homocysteine hydrolase gene (Tanaka *et al.*, *Plant Mol. Biol.* **35**: 981-986, 1997). Our gene expression experiments indicate that OsPN23829 is induced by jasmonic acid which, in addition to having a role in the defense response, inhibits growth processes in many tissues and is active in reproductive development (it is thought to play some role in the formation of flowers, fruit, and seeds; Biochemistry and Molecular Biology of Plants, Buchanan, Gruissem and Jones (eds.), John Wiley & Sons, New York, NY 2002, p. 917). These data suggest that OsPN23829 may be involved in development/plant morphogenesis, and its association with the OsHOS59 may regulate transcriptional events related to these processes. In addition, a metabolic link may exist between the activated methyl cycle reactions mediated by S-adenosyl-L-homocysteine hydrolase and the plant pathogen-induced defense response (Kawalleck *et al.*, *Proc. Natl. Acad. Sci. USA* **89**: 4713-4717, 1992). While no other published evidence points to this conclusion, our gene expression experiments indicate that the gene encoding OsPN23829 is induced by jasmonic acid, which is also a component of plant defense response pathways, and by the fungal pathogen *M. grisea*. It is thus possible that the rice S-adenosyl-L-homocysteine hydrolase OsPN23829 may also have a role in defense against pathogens.

30

The remaining novel proteins found to interact with OsHOS59 include a eukaryotic translation initiation factor 3 large subunit (OsBAB07943) with a putative role in initiation of mRNA translation, a protein similar to latex-abundant protein (OsPN23251), and three proteins similar to *Arabidopsis* proteins of unknown function (OsPN23388, OsPN30858, and a putative PHD-finger protein OsPN23830). The association of these prey proteins with OsHOS59 suggests a role in transcriptional regulation of genes involved in development.

Many of the rice proteins found to interact with the KNOX homeodomain protein OsHOS59 have roles in plant cell cycle/development. This observation corroborates the notion that the previously uncharacterized protein OsHOS59 is involved in transcriptional regulation of development genes. Some of these interactors are newly characterized rice proteins, and their interactions with OsHOS59 represent molecular mechanisms for transcriptional regulation of developmental processes in rice that have not been previously described.

The identification of protein-protein interactions in rice has important commercial applications. Modulation of these interactions may allow control of biological processes mediated by these molecules, resulting in the introduction of desirable traits in genetically engineered plants. The proteins identified in the present Example may be exploited for the development of genetically engineered crops that exhibit desirable changes in plant development. In addition, these proteins may allow the identification of compounds that affect plant development.

Plants can regenerate individual plants through the regeneration of adventitious shoots or adventitious embryos from undifferentiated tissues derived from somatic cells, a process regulated by the interaction of plant hormones such as auxins and cytokinins. In addition to responding to the signals produced by plant hormones, homeobox genes are involved in plant morphogenesis. The regeneration ability of plants is exploited for the production of young plants from cultured shoot and for regenerating transformed plants after the introduction of genes into somatic cell tissues or cultured plant cells. Proposed

applications for homeobox proteins include the control of plant regeneration, differentiation, and growth, processes. For example, genes capable of promoting regeneration of adventitious roots or adventitious shoots from undifferentiated cells or plant tissues would be useful for agricultural applications. In one such application, an
5 *Arabidopsis* gene has been identified encoding a protein with a homeodomain which is involved in differentiation, specifically, it induces adventitious shoots and branching from cultured tissue (see European Patent Application No. EP00946451 EP). In another application, ectopic expression of a plant homeobox gene encoding a transcription factor involved in the metabolism of gibberellic acid and resulting in a delayed flowering
10 phenotype was proposed for the production of genetically modified grasses that exhibit inhibition of flowering, absence of inflorescence, increased production of tillers, delayed heading, and inhibition of the developmental switch from vegetative to generative growth. These modified phenotypes represent agronomically valuable traits in grasses bred for both forage and amenity purposes (see European Patent Application No.
15 EP0109570 EP).

Applications can also be envisioned for the individual proteins identified in this Example. For example, the rice putative eukaryotic translation initiation factor 3 large subunit (OsBAB07943) could be used to identify compounds that inhibit the binding of
20 this plant initiation factor to the cap structure of its mRNAs. Such compounds could function as herbicides. A similar application has been proposed for a plant eukaryotic initiation factor 4E (eIF4E) (Canadian Patent Application No. CA0001412 CA published 6-Jul-2001).

25 Example V

The example describes the identification and characterization of rice proteins that interact at the thylakoid of chloroplasts and other cellular membranes. Specifically, described in this example are newly characterized rice proteins interacting with the rice
14-3-3 protein homolog GF14-c (OsGF14-c) and with Defender Against Apoptotic Death
30 1 (OsDAD1).

The 14-3-3 proteins (reviewed in Muslin and Xing, *Cell Signal* 12(11-12): 703-709, 2000) interact with a variety of regulators of cellular signaling, cell cycle, and apoptosis by binding to their partner proteins. The high potential for specific protein-protein interactions makes these proteins suitable for two-hybrid assays. The 14-3-3 proteins are known to participate in protein complexes within the nucleus and are commonly found in the cytoplasm. Studies using yeast two-hybrid assays have also localized GF14 isoforms to the chloroplast stroma and the stromal side of thylakoid membranes (Sehnke *et al.*, *Plant Physiol.* 122(1): 235-242, 2000). However, the subcellular localization of GF14-c had not been directly assessed to date. Investigation of the protein interactions involving OsGF14-c may lead to the identification of its location within the cell.

OsDAD1 is encoded by the rice homolog of the highly conserved DAD gene, a suppressor of endogenous programmed cell death, or apoptosis, in animals and plants (Apte *et al.*, *FEBS Lett.* 363(3): 304-306, 1995; Gallois *et al.*, *Plant J.* 11(6): 1325-1331, 1997). In support of this role for DAD, expression of a DAD plant homolog has been shown to be down-regulated during flower petal senescence (an example of programmed cell death) and by the plant hormone ethylene, which is associated with a variety of stress responses and developmental processes (Orzaez and Granell, *FEBS Lett.* 404(2-3): 275-278, 1997). While these studies have been conducted with DAD homologs from *Arabidopsis* and pea, the rice DAD1 is not described in the literature. The interaction studies provided below were aimed at further characterizing this protein.

An automated, high-throughput yeast two-hybrid assay technology (as described above) was used to search for rice protein that interacted with the bait proteins OsGF14-c and OsDAD1. The sequences encoding the protein fragments used in the search were then compared by BLAST analysis against databases to determine the sequences of the full-length genes. The proteins found appear to be localized to the thylakoid of chloroplasts, vacuolar membrane and plasma membrane. The results indicate that OsGF14-c is a membrane component in rice. The subset of proteins interacting with OsGF14-c at the thylakoid form a novel chloroplast protein complex involved in the photosynthetic processes. This interaction study also identifies the rice OsDAD1 as a membrane protein, in agreement with previously characterized DAD homologs from

other species. Elucidation of the role of proteins interacting at the thylakoid and other cellular membranes in rice chloroplasts may allow the development of herbicides specifically targeted to disrupting the structure and function of the thylakoid or endomembrane system.

5 This example provides newly characterized rice proteins interacting with the rice 14-3-3 protein homolog GF14-c (OsGF14-c) and with Defender Against Apoptotic Death 1 (OsDAD1). An automated, high-throughput yeast two-hybrid assay technology (provided by Myriad Genetics Inc., Salt Lake City, UT) was used to search for protein interactions with the bait proteins OsGF14-c and OsDAD1. The 14-3-3 proteins
10 (reviewed in Muslin AJ, Xing, *Cell Signal* 12(11-12): 703-709, 2000) interact with a variety of regulators of cellular signaling, cell cycle, and apoptosis by binding to their partner proteins. The high potential for specific protein-protein interactions makes these proteins suitable for two-hybrid assays. The 14-3-3 proteins are known to participate in protein complexes within the nucleus and are commonly found in the cytoplasm. Studies
15 using yeast two-hybrid assays have also localized GF14 isoforms to the chloroplast stroma and the stromal side of thylakoid membranes (Sehnke *et al.*, *Plant Physiol.* 122(1): 235-242, 2000). However, the subcellular localization of GF14-c had not been directly assessed to date. Investigation of the protein interactions involving OsGF14-c may lead to the identification of its location within the cell.

20 OsDAD1 is encoded by the rice homolog of the highly conserved DAD gene, a suppressor of endogenous programmed cell death, or apoptosis, in animals and plants (Apte *et al.*, *FEBS Lett* 363(3): 304-306, 1995; Gallois *et al.*, *Plant J.* 11(6): 1325-1331, 1997). In support of this role for DAD, expression of a DAD plant homolog has been shown to be down-regulated during flower petal senescence (an example of programmed
25 cell death) and by the plant hormone ethylene, which is associated with a variety of stress responses and developmental processes (Orzaez and Granell, *FEBS Lett.* 404(2-3): 275-8, 1997). While these studies have been conducted with DAD homologs from *Arabidopsis* and pea, the rice DAD1 is not described. The interaction studies provided in this example are aimed at characterizing this protein.

30

Results

GF14-c was found to interact with EPSP synthase, an enzyme in the shikimate pathway (OsBAB61062); two enzymes with roles in the Calvin cycle reactions in chloroplasts, a rice chloroplastic aldolase (OsBAA02730) and a the chloroplast enzyme Rubisco (OsRBCL); the Rubisco activase precursor (OsRCAA1); and two rice photosystem proteins, putative 33kDa oxygen-evolving protein of photosystem II (OsPN23059) and photosystem II 10 kDa polypeptide (OsAAB46718). Eight additional interactors for GF14-c are novel rice proteins: a photosystem protein (OsPN23061) similar to barley (*Hordeum vulgare*) photosystem I reaction center subunit II, chloroplast precursor; a protein (OsPN22858) similar to *Arabidopsis thaliana* GTP cyclohydrolase II, an enzyme involved in the biosynthesis of vitamin B riboflavin (a cofactor in the shikimate pathway); a protein (OsPN22874) similar to *A. thaliana* phosphatidylinositol-4-phosphate 5 kinase (PI4P5K), an enzyme involved in signaling events associated with water-stress response in plants; two H⁺-ATPases, similar to *A. thaliana* vacuolar ATP synthase subunit C (OsPN22866) and to barley plasma membrane H⁺-ATPase (OsPN23022); a putative dynamin homolog (OsPN30846) that is likely localized to the chloroplast, as are other plant dynamin family members; and two proteins of unknown function (OsPN29982 and OsPN30974).

OsDAD1 was found to interact with three membrane proteins: rice beta-expansin (OsEXPB2), which is localized to the plasma membrane adjacent to the cell wall; a novel putative phosphate cotransporter (OsPN23053); and the H⁺-ATPase-like protein OsPN23022 that also interacts with GF14-c.

The proteins that interacted with OsGF14-c (14-3-3 protein homolog GF14-c) and OsDAD1 are listed in Tables 21 and 22, respectively, followed by detailed information on each protein and a discussion of the significance of the interactions. A diagram of the interactions is provided in Figure 4. The nucleotide and amino acid sequences of the proteins of the Example are provided in Figure 12.

Nine of the proteins identified represent rice proteins previously uncharacterized. Based on their presumed biological function and on the ability of the prey proteins to specifically interact with the bait proteins OsGF14-c and OsDAD1, it was speculated that OsGF14-c is a membrane component. Based on the results described below, OsGF14-c is presumably localized to the thylakoid of rice chloroplasts and to other cellular membranes. The proteins interacting in the thylakoid are part of a novel protein complex and are involved in the photosynthetic processes occurring in the chloroplasts. Knowledge of the role of proteins interacting at the thylakoid in rice could be exploited for the development of herbicides specifically targeted to disrupting the structure and function of the thylakoid membrane. The interactions found in this study also identify OsDAD1 as a likely membrane component in rice, an observation consistent with previous reports on other animal and plant DAD homologs.

Table 21. Interacting Proteins Identified for OsGF14-c (14-3-3 protein homolog GF14-c).

The Myriad names and the TMRI names of the clones of the proteins used as baits and found as preys are given. Nucleotide/protein sequence accession numbers for the proteins of the Example (or related proteins) are shown in parentheses under the protein name. The bait and prey coordinates (Coord) are the amino acids encoded by the bait fragment(s) used in the search and by the interacting prey clone(s), respectively. The source is the library from which each prey clone was retrieved.

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (source)
BAIT PROTEIN :			
OsGF14-c PN12464	<i>O. sativa</i> 14-3-3 Protein Homolog GF14-c (U65957)	1-257#	
INTERACTORS :			
OsBAB61062 PN22844	<i>O. sativa</i> 3-Phosphoshikimate 1-carboxyvinyltransferase (a.k.a. EPSP Synthase) (AB052962; BAB61062.1)	1-150	463-511 (input trait)
OsPN22858	Novel Protein 22858, Fragment, similar to <i>Arabidopsis</i> GTP Cyclohydrolase II (BAB09512.1; e=0)	1-150	27-154 (input trait)
OsPN22874	Novel Protein 22874, Fragment, similar to <i>Arabidopsis</i> Putative Phosphatidylinositol-4-phosphate 5-kinase (NP_187603.1; 4e ⁻¹⁸)	1-150	1-88 (input trait)
OsBAA02730 PN22832 (Contig4280.fasta.Con tig1)	<i>O. sativa</i> Fructose-Bisphosphate Aldolase, Chloroplast Precursor (Q40677)	1-150	206-269 (input trait)

OsRBCL PN23426	<i>O. sativa</i> Chloroplast Ribulose Bisphosphate Carboxylase, Large Chain (D00207; P12089)	1-150	287-462 (input trait)
OsRCAA1 PN19842	<i>O. sativa</i> Ribulose Bisphosphate Carboxylase/Oxygenase Activase, Large Isoform A1 (AB034698, BAA97583)	1-150	68-210 (input trait)
OsPN22866 (Contig388.fasta.Contig2)	Novel Protein PN22866, Fragment, Similar to <i>A. Thaliana</i> Vacuolar ATP Synthase Subunit C (V-ATPase C subunit) (Vacuolar proton pump C subunit) (Q9SDS7; e ⁻¹⁵²)	1-150	95-305 (input trait)
OsPN23022§	Novel Protein PN23022, Fragment, similar to <i>H. Vulgare</i> Plasma Membrane H ⁺ -ATPase (CAC50884; e=0.0)	1-150	149-285 (input trait)
OsPN23061 (Contig3864.fasta.Contig1)	Hypothetical Protein OsContig3864, Similar to <i>H. vulgare</i> Photosystem I Reaction Center Subunit II, Chloroplast Precursor (P36213; 6e ⁻⁸⁷)	1-150	94-203 (input trait)
OsPN23059 (Contig4331.fasta.Contig1)	OsContig4331, <i>O. sativa</i> Putative 33kDa Oxygen-Evolving Protein of Photosystem II (BAB64069)	1-150	193-333 90-169 (input trait)
OsAAB46718 PN22840 (FL_R01_003_H20.g.1a.Sp6a TMRI)	<i>O. sativa</i> Photosystem II 10 kDa Polypeptide (U86018; T04177)	1-150	82-126 (input trait)
OsPN29982	Novel Protein PN29982	1-150	201-300 (input trait)
OsPN30846	Novel Protein PN30846	1-150	1-266 (input trait)
OsPN30974	Novel Protein PN30974	1-150	38-178 (input trait)

NOTE: Interactions of GF14-c with the maize transcription factor Viviparous-1 (ZmVP1) and with Em binding protein (EmBp) are also reported in the literature (Schultz *et al.*, *Plant Cell* 10(5): 837-47, 1998).

Self-activating clone, i.e., it activates the reporter genes in the two-hybrid system in the absence of a prey protein, and thus it was not used in the search.

\$ A prey clone of OsPN23022 also interacts with a clone of Defender Against Apoptotic Death 1 (OsDAD1) used as a bait, and the bait OsDAD1 interacts with Beta-Expansin EXPB2 (OsEXPB2) and with Novel Protein 23053, Fragment, Similar to *Arabidopsis* Putative Na⁺-Dependent Inorganic Phosphate Cotransporter (OsPN23053). These interactions are shown in Table 22 below.

Table 22. Interacting Proteins Identified for OsDAD1 (Defender Against Apoptotic Death 1).

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (source)
BAIT PROTEIN :			
OsDAD1 PN20251	<i>O. sativa</i> Defender Against Apoptotic Death 1 (D89727; BAA24104)		
INTERACTORS:			
OsPN23022	Novel Protein PN23022, Fragment, similar to <i>H. Vulgare</i> Plasma Membrane H ⁺ -ATPase (CAC50884; e=0.0)	30-115	37-371 (input trait)
OsPN23053	Novel Protein 23053, Fragment, Similar to <i>Arabidopsis</i> Putative Na ⁺ -Dependent Inorganic Phosphate Cotransporter (NP_181341.1; e ⁻¹⁰⁵)	30-115	2x 1-180 (input trait)
OsEXPB2 PN19902	Beta-Expansin EXPB2 (U95968; AAB61710)	1-115	80-207 (input trait)
		30-115	183-261 2x 80-218 (input trait)

Two-hybrid system using Os GF14-c as bait

GF14-c (GenBank Accession #U65957) is a 256-amino acid protein that has been reported to interact with site-specific DNA-binding proteins (*i.e.*, basic leucine zipper factor EmBP1) and tissue-specific regulatory factors (*i.e.*, viviparous-1; VP-1) (Schultz *et al.*, *Plant Cell* 10(5): 837-847, 1998). It may act to form complexes with EmBP1 and VP-1 to mediate gene expression. The 14-3-3 proteins are found in virtually every eukaryotic organism and tissue and usually consist, in any given organism, of multiple protein isoforms (De Lille *et al.*, *Plant Physiol.* 126(1): 35-38, 2001). They are thought to act as molecular scaffolds or chaperones and to regulate the cytoplasmic and nuclear localization of proteins with which they interact by regulating their nuclear import/export (Zilliacus *et al.*, *Mol. Endocrinol.* 15(4): 501-511, 2001; reviewed by Muslin and Xing, *Cell Signal* 12(11-12): 703-709, 2000). The 14-3-3 proteins bind to a multitude of functionally diverse regulatory proteins involved in cellular signaling pathways, cell cycling, and apoptosis. In plants, enzymes under the control of 14-3-3 proteins include starch synthase, Glu synthase, F1 ATP synthase, ascorbate peroxidase, and caffeoyl methyl transferase, plasmamembrane H⁺-ATPase, light- and substrate-regulated metabolic enzymes of the nitrogen and carbon assimilation pathways, and those involved in transcriptional regulation such as the G-box complex and core transcription factors TBP, TFIIB, and EmBP. However, the specific 14-3-3 isoforms required by each of

these pathways have not been fully characterized (De Lille *et al.*, *supra*). The 14-3-3 proteins have previously been detected as participants in protein complexes within the nucleus (Bihn *et al.*, *Plant J.* 12(6): 1439-1445, 1997; Imhof A, Wolffe, *Biochemistry* 38(40): 13085-13093, 1999; Zilliacus *et al.*, *supra*), in the cytoplasm, and mitochondria (De Lille *et al.*, *supra*). Plant 14-3-3 proteins have also been localized to the chloroplast stroma and the stromal side of thylakoid membranes (Sehnke *et al.*, *supra*). However, subcellular localization of GF14-c has not been directly assessed and thus its location within the cell is yet to be precisely defined.

10 Analysis of the amino acid sequence of GF14-c identified a cAMP- and GMP-dependent phosphorylation site at amino acids 107 to 110, six protein kinase C phosphorylation sites (amino acids 10 to 12, 29 to 31, 56 to 61, 29 to 31, 59 to 61, and 74 to 76), three casein kinase II phosphorylation sites (amino acids 110 to 113, 120 to 123, and 177 to 180), an N-myristoylation site (amino acids 9 to 14), and two amidation sites
15 (amino acids 77 to 80 and 105 to 108). The bait fragment used in this search encodes amino acids 1 to 150 of GF14-c. A BLAST analysis comparing the nucleotide sequence of GF14-c against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS009195_at (e^{-48} expectation value) as the closest match. Gene expression experiments indicated that this gene is not specifically expressed in several different
20 tissue types and is not specifically induced by a broad range of stresses, herbicides and applied hormones.

 The bait protein encoding amino acids 1 to 150 of GF14-c was found to interact with *O. sativa* 3-phosphoshikimate 1-carboxyvinyltransferase (a.k.a. EPSP Synthase)
25 (OsBAB61062). OsBAB61062 is a 511-amino acid protein that contains an EPSP synthase signature 1 site (amino acids 162 to 176), an EPSP signature 2 site (amino acids 423 to 441), and it is alanine-rich at the N-terminus. A BLAST analysis of the amino acid sequence of OsBAB61062 determined that this protein is the rice 3-phosphoshikimate 1-carboxyvinyltransferase (also commonly referred to as EPSP
30 synthase) (GenBank Accession No. BAB61062.1, 83.9% identity, $e=0.0$). This 511-amino acid enzyme is located in the chloroplasts where it catalyzes an essential step in

aromatic amino acid synthesis, referred to as the shikimate pathway. Because EPSP synthase is essential to algae, higher plants, bacteria, and fungi, but not present in mammals, this enzyme is a useful herbicide and antimicrobial target.

A BLAST analysis comparing the nucleotide sequence of EPSP synthase against
5 TMRI's GeneChip® Rice Genome Array sequence database identified probeset
OS020639.1_at (e^{-156} expectation value) as the closest match. Gene expression
experiments indicated that this gene is induced by jasmonic acid, a plant hormone
involved in signal transduction events associated with a plant's stress response, and by *M.*
10 *grisea*, the fungus that causes rice blast disease. The gene is repressed under drought
conditions.

The bait protein encoding amino acids 1 to 150 of GF14-c was found to interact
with protein 22858, a fragment which is similar to *A. thaliana* GTP cyclohydrolase II
(OsPN22858). This prey clone of OsPN22858 is a 460-amino acid protein fragment with
15 a transmembrane region spanning amino acids 182 to 198 and a possible cleavage site
between amino acids 24 and 25, although no N-terminal signal peptide is present. A
BLAST analysis of OsPN22858 determined that its amino acid sequence most nearly
matches that of GTP cyclohydrolase II; 3,4-dihydroxy-2-butanone-4-phosphate synthase
from *A. thaliana* (GenBank Accession # BAB09512.1, 74.4% identity, $e=0$). GTP
20 cyclohydrolase II catalyzes the first committed reaction in the biosynthesis of the B
vitamin riboflavin (Ritz *et al.*, *J. Biol. Chem.* 276(25): 22273-22277, 2001).

A BLAST analysis comparing the nucleotide sequence of Novel Protein 22858
against TMRI's GeneChip® Rice Genome Array sequence database identified
OS015318_s_at ($5e^{-10}$ expectation value) as the closest match. The expectation value is
25 too low for this probeset to be a reliable indicator of the gene expression of this GTP
cyclohydrolase.

The bait protein encoding amino acids 1 to 150 of GF14-c was found to interact
with Protein 22874, a fragment that is similar to *A. thaliana* putative
30 phosphatidylinositol-4-phosphate 5-kinase (OsPN22874). A BLAST analysis of
OsPN22874 determined that its 89-amino acid sequence most nearly matches that of

phosphatidylinositol-4-phosphate 5-kinase (PI4P5K) from *A. thaliana* (GenBank Accession No. NP_187603.1, 65.5% identity, $4e^{-18}$). PI4P5K is an enzyme that plays a well-defined role in many signaling events in many species, including the endoplasmic reticulum (ER) stress response in plants (Shank *et al.*, *Plant Physiol.* **126**(1): 267-277, 2001). Animal and yeast PI4P5K phosphorylates phosphatidylinositol-4-phosphate to produce phosphatidylinositol-4,5-bisphosphate as a precursor of two second messengers, inositol-1,4,5-triphosphate and diacylglycerol, and as a regulator of many cellular proteins involved in signal transduction and cytoskeletal organization (reviewed in Mikami *et al.*, *Plant J.* **15**(4): 563-568, 1998). Mikami *et al.* identified a full-length cDNA clone encoding a PI4P5K protein in *A. thaliana* whose mRNA expression is induced by treatment of the plant with drought, salt and abscisic acid, suggesting that this protein is involved in water-stress signal transduction (Mikami *et al.*, *supra*). Elge *et al.* report that *A. thaliana* PI4P5K is expressed predominantly in vascular tissues of leaves, flowers and roots, namely in cells of the lateral meristem, i.e., the procambium (Elge *et al.*, *Plant J.* **26**(6): 561-571, 2001).

The bait protein encoding amino acids 1 to 150 of GF14-c was also found to interact with *O. sativa* fructose-bisphosphate aldolase, a chloroplast precursor (OsBAA02730). OsBAA02730 (GenBank Accession No. Q40677) is a 388-amino acid protein that includes a fructose-bisphosphate aldolase class-I active site (amino acids 44 and 388), as determined by analysis of the amino acid sequence ($8.5e^{-228}$). A BLAST analysis of the amino acid sequence of OsBAA02730 indicated that this protein is the rice fructose-bisphosphate aldolase, chloroplast precursor (GenBank Accession No. Q40677). The gene encoding chloroplastic aldolase was isolated along with that encoding the cytoplasmic form of the enzyme (Tsutsumi *et al.*, *Gene* **141**(2): 215-220, 1994). The chloroplastic aldolase is encoded at a single locus, while the cytoplasmic form is distributed between three loci on the genome. Aldolases are present in higher plants as two isoforms, the cytosolic and the chloroplastic types. The cytoplasmic form is highly conserved among plants and appears to be regulated through a Ca^{2+} -mediated protein kinase/phosphatase pathway (Nakamura *et al.*, *Plant Mol. Biol.* **30**(2): 381-385, 1996). This enzyme is thought to have a role in the fruit ripening process (Schwab *et al.*, *Phytochemistry* **56**(5): 407-415, 2001). The chloroplastic enzyme is involved in two

major sugar phosphate metabolic pathways of green chloroplasts: the C3 photosynthetic carbon reaction cycle (Calvin cycle) and reactions of the starch biosynthetic pathway. In both cases, aldolase catalyzes the formation of fructose 1,6-biphosphate from dihydroxyacetone 3-phosphate and glyceraldehyde 3-phosphate. These topics are reviewed by Michelis *et al.* (*Plant Mol. Biol.* 44(4): 487-498, 2000), who also identified a 44-kDa heat-induced isoform of the fructose-bisphosphate aldolase in oat chloroplast, confirming its localization to the thylakoid membrane and suggesting that this enzyme is not embedded but rather tends to adhere to the chloroplast membranes. Similar heat-induced thylakoid-associated aldolase homologues were found in other plant species.

A BLAST analysis comparing the nucleotide sequence of the aldolase protein against TMRI's GeneChip[®] Rice Genome Array sequence database identified probeset OS006916.1_at (e^{-156} expectation value) as the closest match. Our gene expression experiments indicate that this gene is down-regulated by jasmonic acid and drought.

In addition, the bait protein encoding amino acids 1 to 150 of GF14-c was found to interact with *O. sativa* ribulose bisphosphate carboxylase large chain precursor (RuBisCO Large Subunit) (OsRBCL). A BLAST analysis of the amino acid sequence of OsRBCL determined that this protein is the rice chloroplast ribulose bisphosphate carboxylase, large chain precursor ((RuBP carboxylase/oxygenase, also called Rubisco for short) (GenBank Accession No. P12089). Rubisco is a 477-amino acid protein present in the chloroplast of higher plants, with an active site in position 196-204. The chloroplast RuBP carboxylase/oxygenase is part of the CO₂-fixing multienzyme complexes bound to the thylakoid membrane (Suss *et al.*, *Proc. Natl. Acad. Sci. USA* 90(12): 5514-5518, 1993) with roles in the Calvin cycle reactions that occur in the stroma of the chloroplast during photosynthesis. The starting and ending compound in the Calvin cycle is the five-carbon sugar ribulose 1,5-biphosphate (RuBP). As its name indicates, RuBP carboxylase/oxygenase catalyzes two types of reactions that involve RuBP. In the presence of high carbon dioxide and low oxygen concentrations, the carboxylase activity of Rubisco is favored and the enzyme catalyzes the initial reaction in the Calvin cycle, the carboxylation of RuBP, leading to the formation of 3-phosphoglyceric acid (PGA). However, in the presence of low carbon dioxide and high

oxygen concentrations, oxygen competes with carbon dioxide as a substrate for Rubisco and the enzyme's oxygenase activity also occurs, resulting in condensation of oxygen with RuBP to form 3-phosphoglycerate and phosphoglycolate. Rubisco is the world's most abundant enzyme, accounting for as much as 40 percent of total soluble protein in leaves (these topics are discussed in Raven, Evert, and Eichhorn, Biology of Plants (6th Ed.), W.H. Freeman, New York, NY 1999).

A BLAST analysis comparing the nucleotide sequence of the Rubisco protein against TMRI's GeneChip[®] Rice Genome Array sequence database identified probeset OS000296_s_at ($e=0$ expectation value) as the closest match. Gene expression experiments indicated that this gene is down-regulated by BAP, 2,4-D, BL2, jasmonic acid, gibberellin, and abscisic acid. The gene is up-regulated under osmotic stress conditions.

The bait protein encoding amino acids 1 to 150 of GF14-c was found to interact with *O. sativa* ribulose biphosphate carboxylase/oxygenase activase, large isoform A1 (OsRCAA1). A BLAST analysis of the amino acid sequence of OsRCAA1 determined that this 466-amino acid protein is the rice Rubisco activase large isoform precursor (GenBank Accession No. BAA97583). It contains two active sites (amino acid 31 to 38 and 156 to 163). Rubisco activase is an AAA+ (ATPases associated with a variety of cellular activities) protein that facilitates the ATP-dependent removal of sugar phosphates from Rubisco active sites. This action frees the active site of Rubisco for spontaneous carbamylation by CO₂ and metal binding, prerequisites for activity (reviewed in Salvucci *et al.*, *Plant Physiol.* 127(3): 1053-1064, 2001; Salvucci and Ogren, *Photosynthesis Res.* 47: (1) 1-11, 1996).

25

The bait protein encoding amino acids 1 to 150 of GF14-c was found to interact with protein PN22866, a fragment similar to *A. thaliana* vacuolar ATP synthase subunit C (V-ATPase C subunit) (vacuolar proton pump C subunit) (OsPN22866). OsPN22866 is a 408-amino acid protein fragment. Its amino acid sequence most nearly matches that of *A. thaliana* Vacuolar ATP synthase subunit C (V-ATPase C subunit) (Vacuolar proton pump C subunit) (Q9SDS7, 72.7% identity, e^{-152}), as determined by BLAST analysis.

30

The H⁺-translocating ATPases (H⁺-ATPase, V-ATPase) are multi-subunit enzymes that function as essential proton pumps in eukaryotes. The catalytic site of human V-ATPase consists of a hexamer of three A subunits and three B subunits that bind and hydrolyze ATP and are regulated by accessory subunits C, D and E (van Hille *et al.*, *Biochem*
5 *Biophys Res. Commun.* **197**(1): 15-21, 1993).

ATPases are essential cellular energy converters that transduce the chemical energy of ATP hydrolysis from transmembrane ionic electrochemical potential differences. The plant ATPases are present in chloroplasts, mitochondria and vacuoles. In vacuoles, ATPases regulate the contents and volume of vacuoles, which depends on
10 the coordinated activities of transporters and channels located in the tonoplast (vacuolar membrane). The V-ATPase uses the energy released during cleavage of the phosphate group of cytosolic ATP to pump protons into the vacuolar lumen, thereby creating an electrochemical H⁺-gradient that is the driving force for transport of ions and metabolites. Thus V-ATPase is important as a 'house-keeping' and as a stress response enzyme.
15 Expression of V-ATPase has been shown to be highly regulated depending on metabolic conditions. The V-ATPase consists of several polypeptide subunits that are located in two major domains, a membrane peripheral domain (V₁) and a membrane integral domain (V_o). Subunit C is a highly hydrophobic protein containing four membrane-spanning domains. The function of subunit C is unknown, although it is suggested to be
20 directly involved in H⁺ transport and might be involved in stabilization of V₁. The structure, function and regulation of the plant V-ATPase are reviewed in Ratajczak R., *Biochim Biophys Acta* **1465**(1-2): 17-36, 2000.

The bait protein encoding amino acids 1 to 150 of GF14-c was also found to
25 interact with protein PN23022, a fragment similar to *H. Vulgare* plasma membrane H⁺-ATPase (OsPN23022). Protein PN23022 is a 534-amino acid fragment that includes seven transmembrane domains (amino acids 170 to 186, 202 to 218, 226 to 242, 266 to 282, 308 to 324, 337 to 353, and 373 to 389), as predicted by analysis of its amino acid sequence. A BLAST analysis of the amino acid sequence of OsPN23022 determined that
30 this protein is similar to *H. vulgare* plasma membrane H⁺-ATPase (GenBank Accession No. CAC50884; 88.2% identity, e=0 expectation value), an enzyme that translocates

protons into intracellular organelles or across the plasma membrane of eukaryotic cells. A BLAST analysis comparing the nucleotide sequence of Novel protein PN23022 against TMRI's GeneChip® Rice Genome Array sequence database identified OS000972_f_at (e^{-11} expectation value) as the closest match. The expectation value is too low for this probeset to be a reliable indicator of the gene expression of this ATPase. OsPN23022 was also found to interact with Defender Against Apoptotic Death 1 (OsDAD1) (see Table 22).

The bait protein encoding amino acids 1 to 150 of GF14-c was found to interact with protein OsContig3864, which is similar to *H. vulgare* photosystem I reaction center subunit II, chloroplast precursor (OsPN23061). Analysis of the OsContig3864 amino acid sequence predicted that it is a 203-amino acid protein containing a possible cleavage site between amino acids 21 and 22, although there appears to be no N-terminal signal peptide. A BLAST analysis determined that the OsContig3864 clone has an amino acid sequence that most nearly matches that of *H. vulgare* photosystem I reaction center subunit II, chloroplast precursor (Photosystem I 20 kDa subunit) (PSI-D) (GenBank Accession No. P36213, 80% identity, $3e^{-86}$). The photosystems (photosystems I and II) are large multi-subunit protein complexes embedded into the photosynthetic thylakoid membrane. They operate in series and catalyze the primary step in oxygenic photosynthesis, the light-induced charge separation process by which light energy from the sun is converted to carbon dioxide and carbohydrates in plants and cyanobacteria. Photosystem I catalyzes the light-induced electron transfer from plastocyanin/cytochrome c_6 on the luminal side of the membrane (inside the thylakoids) to ferredoxin/flavodoxin at the stromal side by a chain of electron carriers (reviewed in Fromme *et al.*, *Biochim. Biophys Acta* **1507**(1-3): 5-31, 2001).

A BLAST analysis comparing the nucleotide sequence of OsContig3864 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS000721_at ($e=0$ expectation value) as the closest match. Gene expression experiments indicated that this gene is not specifically expressed in several different plant tissue types and is not specifically induced by a broad range of stresses, herbicides and applied hormones.

The bait protein encoding amino acids 1 to 150 of GF14-c was also found to interact with OsContig4331, an *O. Sativa* putative 33kDa oxygen-evolving protein of photosystem II (OsPN23059). The two prey clones retrieved from the input trait library encode amino acids 193 to 333 and 90 to 169 of OsContig4331. These clones are non-overlapping, suggesting that multiple GF14-c-binding sites exist within OsContig4331. Analysis of the OsContig4331 protein sequence predicted that it codes for a 333-amino acid protein. The analysis also indicated that OsContig 4331 contains a possible cleavage site between amino acids 37 and 38, although no N-terminal signal peptide is evident. A BLAST analysis of the OsContig 4331 amino acid sequence determined that this protein is the rice putative 33kDa oxygen-evolving protein of photosystem II (GenBank Accession No. BAB64069, 90.6% identity, e^{-169}). Photosystem II uses photooxidation to convert water to molecular oxygen, thereby releasing electrons into the photosynthetic electron transfer chain.

A BLAST analysis comparing the nucleotide sequence of OsContig4331, rice Photosystem I Reaction Center Subunit II Precursor against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS000372_at ($e=0$ expectation value) as the closest match. Our gene expression experiments indicate that this gene is down-regulated during cold stress.

The bait protein encoding amino acids 1 to 150 of GF14-c was also found to interact with *O. Sativa* photosystem II 10 kDa polypeptide (OSAAB46718). OSAAB46718 is a 126-amino acid protein fragment that includes a predicted transmembrane domain (amino acids 102 to 118). A BLAST analysis against the Genpept database revealed that OSAAB46718 is the *Oryza sativa* photosystem II 10kDa polypeptide (GenBank Accession No. T04177, 91.2% identity, $2e^{-61}$).

The bait protein encoding amino acids 1 to 150 of GF14-c was also found to interact with protein PN29982 (OsPN29982). The 300-amino acid sequence of the protein OsPN29982 most nearly matches that of a putative protein of unknown function from *A. thaliana* (GenBank Accession No. NP_196688.1, 47% identity, $3e^{-054}$), as determined by BLAST analysis. The second best match was CHICK LIM/homeobox

protein Lhx1 (Homeobox protein LIM-1) (GenBank Accession No. P53411, 28% identity, $e=0.002$). Based on the homeobox domain, this interaction may be similar to 14-3-3 protein interactions with transcription factors like VP1.

5 The bait protein encoding amino acids 1 to 150 of GF14-c was also found to interact with protein PN30846 (OsPN30846). A BLAST analysis of protein OsPN30846 determined that its 266-amino acid sequence most nearly matches that of dynamin homolog from the leguminous plant *Astragalus sinicus* (GenBank Accession No. AAF19398.1, 70.6% identity, $2e^{-99}$). Since the discovery of the GTP-binding dynamin in
10 rat brain, dynamin-like proteins have been isolated from various organisms and tissues and shown to be involved in diverse and seemingly unrelated biological processes. Many different isoforms of dynamin-like proteins have been identified in plant cells, and these plant homologs can be grouped into several subfamilies, such as G68/ADL1, ADL2 and ADL3, based on their amino acid sequence similarity (reviewed in Kim *et al.*, *Plant*
15 *Physiol.* **127**(3): 1243-1255, 2001). The biological roles have been characterized for a few of these plant dynamin-like proteins. The dynamin-like protein ADL1 from *Arabidopsis* has been shown to be localized to and to be involved in biogenesis of the thylakoid membranes of chloroplasts (Park *et al.*, *EMBO J.* **17**(4): 859-867, 1998). Another *Arabidopsis* dynamin-like protein, ADL2, is targeted to the plastid, and its
20 recombinant form expressed in *E. coli* binds specifically to phosphatidylinositol 4-phosphate through the pleckstrin homology (PH) domain present in ADL2 (Kim *et al.*, *supra*). Based on the similarity between the biochemical properties of ADL2 and those of dynamin and other related proteins, ADL2 may be involved in vesicle formation at the chloroplast envelope membrane.

25 The bait protein encoding amino acids 1 to 150 of GF14-c was also found to interact with protein PN30974 (OsPN30974). A BLAST analysis of the novel protein OsPN30974 determined that its 476-amino acid sequence most nearly matches that of an *Arabidopsis* hypothetical protein of unknown function (GenBank Accession No. NP_173623.1, 49% identity, e^{-137}). The next 13 best hits with an expectation value $<e^{-15}$
30 are all *Arabidopsis* or rice proteins of unknown function annotated in the public domain.

Two-hybrid system using OsDAD1 as bait

A second bait protein, namely *O. sativa* Defender Against Apoptotic Death 1 (OsDAD1), was used to identify interactors. OsDAD1 (GenBank Accession No. BAA24104) is a 114-amino acid protein that includes three predicted transmembrane domains (amino acids 33 to 49, 59 to 75, and 94 to 110). DAD1 is a suppressor of programmed cell death, or apoptosis, a process in which unwanted cells are eliminated during growth and development. DAD is a highly conserved protein with homologs identified in animals and plants (Apte *et al.*, *FEBS Lett* 363(3): 304-306, 1995; Gallois *et al.*, *Plant J.* 11(6): 1325-1331, 1997). Dysfunction and down-regulation of this gene has been linked to programmed cell death in these organisms (Lindholm *et al.*, *Mech. Dev.* 93(1-2): 169-173, 2000). DAD1 is an essential subunit of the oligosaccharyltransferase that is located in the ER membrane (Lindholm *et al.*, *supra*). DAD1 expression declines dramatically upon flower anthesis disappearance in senescent petals and is down-regulated by the plant hormone ethylene (Orzaez and Granell, *FEBS Lett.* 404(2-3): 275-278, 1997), which is involved in a variety of stress responses and developmental processes including petal senescence (Shibuya *et al.*, *J. Exp. Bot.* 51(353): 2067-2073, 2000), cell elongation, cell fate patterning in the root epidermis, and fruit ripening (Ecker, J.R., *Science* 268(5211): 667-675, 1995).

Two clones, encoding amino acids 1-115 and 30-115 of OsDAD1, were used as baits in this Example.

OsDAD1 was found to interact with protein 23053, a fragment which is similar to *Arabidopsis* putative Na⁺-dependent inorganic phosphate cotransporter (OsPN23053). OsPN23053 is a protein fragment; however, its available 379-amino acid sequence contains five predicted transmembrane regions (amino acids 100 to 116, 118 to 134, 226 to 242, 259 to 275, and 324 to 340) and a cleavable signal peptide (amino acids 1 to 46). A BLAST analysis determined that OsPN23053 is similar to an *Arabidopsis* putative Na⁺-dependent inorganic phosphate cotransporter (GenBank Accession No. NP_181341.1, 55.4% identity, e⁻¹⁰⁵). In mammals, Na⁺-dependent inorganic phosphate cotransporter is present in neuronal synaptic vesicles and endocrine synaptic-like

microvesicles as a vesicular glutamate transporter and is responsible for storage of glutamate, the major excitatory neurotransmitter in the mammalian central nervous system (CNS) (Takamori *et al.*, *Nature* 407(6801): 189-194, 2000). At least two isoforms of Na⁺-dependent inorganic phosphate cotransporter exist (Takamori *et al.*,
5 *supra*; Aihara *et al.*, *J. Neurochem.* 74(6): 2622-2625, 2000) and are expressed in pancreas and brain (Hayashi *et al.*, *J. Biol. Chem.* 276(46): 43400-43406, 2001; Fujiyama *et al.*, *J. Comp. Neurol.* 435(3): 379-387, 2001). OsPN23053 is the first of a family of Na⁺-dependent inorganic phosphate cotransporters to be discovered in rice. Plants utilize glutamate in important biological processes including protein synthesis and glutamate-
10 mediated signaling (Lacombe *et al.*, *Science* 292(5521): 1486-1487, 2001). The formation of glutamate from glutamine during nitrogen recycling (Singh *et al.*, *J. Plant. Physiol.* 153(3-4): 316-323, 1998) and the control of nitrogen assimilatory pathways by light-signaling (Oliveira *et al.*, *Braz. J. Med. Biol. Res.* 34(5): 567-575, 2001) in plants suggest a link between glutamate formation and light-signal transduction.

15 OsDAD1 was found to interact with beta-expansin EXPB2 (OsEXPB2). A BLAST analysis of the amino acid sequence of OsEXPB2 determined that this protein is rice beta-expansin (GenBank Accession No. AAB61710, 99.6% identity, e⁻¹⁵⁶). Expansins promote cell wall extension in plants. Shcherban *et al.* isolated two cDNA clones from cucumber that encode expansins with signal peptides predicted to direct
20 protein secretion to the cell wall Shcherban *et al.*, *Proc. Natl. Acad. Sci. USA* 92(20): 9245-9249, 1995). These authors identified at least four distinct expansin cDNAs in rice and at least six in *Arabidopsis* from collections of anonymous cDNAs (Expressed Sequence Tags). They determined that expansins are highly conserved in size and sequence and suggest that this multigene family formed before the evolutionary
25 divergence of monocotyledons and dicotyledons. Their analyses indicate no similarities to known functional domains that might account for the action of expansins on wall extension, though a series of highly conserved tryptophans may mediate expansin binding to cellulose or other glycans.

30

Summary

The thylakoid membrane of the chloroplasts contains the photosynthetic pigments, reaction centres and electron transport chains associated with photosynthesis. Localization of OsGF14-c to this site is consistent with the interactions of OsGF14-c with the photosystem proteins of this Example. The photosystems (photosystems I and II) are large multi-subunit protein complexes embedded in the thylakoid membrane. As part of a larger group of protein-pigment complexes, the photosynthetic reaction centers, they catalyze the light-induced charge separation associated with photosynthesis. Both photosystems use the energy of photons from sunlight to translocate electrons across the thylakoid membrane via a chain of electron carriers. The electron transfer processes are coupled to a build-up of a difference in proton concentration across the thylakoid membrane. The resulting electrochemical membrane potential drives the synthesis of ATP, which is used to reduce CO₂ to carbohydrates in the subsequent dark reactions. OsGF14-c is found to interact with OsContig3864, similar to photosystem I reaction center subunit II, chloroplast precursor, with OsContig4331, the rice putative 33kDa oxygen-evolving protein of photosystem II, and with rice photosystem II 10 kDa polypeptide. The validity of these interactions is supported by results in a report by Sehnke *et al.* (*Plant Physiol* 122(1): 235-242, 2000) who used yeast two-hybrid technology to identify an interaction between a plant 14-3-3 protein and another photosystem I subunit protein, *A. thaliana* photosystem I N-subunit At pPSI-N. The interactions of OsGF14-c with OsPN23061 (OsContig3864), OsPN23059 (OsContig4331), and OsAAB46718 (photosystem II 10 kDa polypeptide) suggest that OsGF14-c has a role in coupling the physical contact between proteins in or on the periphery of thylakoid membranes.

25

Given the interactions of OsGF14-c and components of the chloroplast photosystem, some of the other proteins found to interact with OsGF14-c in this study are likely to be localized to the chloroplast as well, and they are possibly co-located to the thylakoid membrane as interaction complexes. For example, OsGF14-c interacts with EPSP synthase (OsBAB61062), a shikimate pathway enzyme located in the chloroplast, where aromatic amino acid synthesis initiates. It is interesting to note that an enzyme in

30

the shikimate pathway requires a flavin as a cofactor (Bornemann et al., Biochemistry 35(30): 9907-9916, 1996) and that OsGF14-c also interacts with OsPN22858, a novel protein fragment similar to *A. thaliana* GTP cyclohydrolase II. GTP cyclohydrolase II participates in the biosynthesis of the B vitamin riboflavin, which is a cofactor for enzymes functioning in the shikimate pathway. The interactions of these proteins with OsGF14-c may keep key proteins of the shikimate pathway in close proximity in or at the thylakoid. The interactions of OsGF14-c with chloroplastic aldolase (OsBAA02730), an enzyme shown to be localized to the thylakoid membrane and involved in the sugar phosphate metabolic pathway of chloroplasts, and with the Calvin cycle enzyme Rubisco (OsRBCL) and Rubisco activase large isoform precursor (OsRCAA1) further support localization of OsGF14-c and these interactors to the thylakoid membrane. Previous reports have identified a fructose-bisphosphate aldolase isoform at the thylakoid membrane in oat chloroplasts (Michelis *et al.*, *supra*).

In addition, a novel interactor identified for OsGF14-c is a putative dynamin homolog (OsPN30846). Plant dynamin-like proteins have been localized to the thylakoid and envelope membranes of chloroplasts Park *et al.*, *EMBO J.* 17(4): 859-867, 1998; Kim *et al.*, *Plant Physiol.* 127(3): 1243-1255, 2001). Thus it is likely that this rice dynamin homolog is a membrane protein that resides in the chloroplast. This and the fact that other interactors identified for OsGF14-c are present in the thylakoid of chloroplasts substantiates the notion that the 14-3-3 protein functions as a component of the thylakoid or envelope membrane of chloroplasts. In further support of this hypothesis, a recombinant *Arabidopsis* dynamin-like protein member of the ADL2 subfamily binds specifically to phosphatidylinositol 4-phosphate. The interactions between dynamins and phosphoinositides documented in the literature (reviewed in Kim *et al.*, *supra*) are consistent with the concomitant presence of the dynamin-like protein OsPN30846 and the phosphatidylinositol-4-phosphate 5-kinase OsPN22874 (rice PI4P5K), both interacting with OsGF14-c, at the thylakoid. We speculate that the interactors described above are part of a protein complex involved in the photosynthetic processes at the thylakoid membrane.

In addition to components of the chloroplast thylakoid, OsGF14-c was found to interact with proteins similar to a plasma membrane H^+ -ATPase (OsPN23022) and to a vacuolar ATPase (OsPN22866), which suggests that OsGF14-c is also present in plasma and vacuolar membranes. The interactions of OsGF14-c with the ATPases may represent 14-3-3 regulation of the plant turgor pressure. This hypothesis is corroborated by reports of 14-3-3 proteins accomplishing this function via regulation of at least one form of a plasma membrane H^+ ATPase (reviewed in DeLille *et al.*, *Plant Physiol.* **126**(1): 35-38, 2001). The interaction of the vacuolar ATPase with OsGF14-c may occur in the vacuolar membrane, but also in membranes of the ER, Golgi bodies, coated vesicles, and provacuoles.

The biological significance of the interaction of OsGF14-c with the novel protein OsPN22874 (rice PI4P5K) may be defined based on functional homology with *A. thaliana* PI4P5K, which is induced under water-stress conditions and is expressed in leaves. Given the interaction of OsGF14-c with components of the thylakoid and vacuolar membranes, the rice PIP5K may be located in the chloroplast but it may also reside at the vacuole, with the vacuolar ATPase. In either case, the rice PIP5K may direct synthesis of molecules involved in kinase signaling events associated with chloroplast protection or vacuole size regulation under abiotic stress.

Two additional interactors, OsPN29982 and OsPN30974, found for OsGF14-c are proteins of unknown function. Nevertheless, because 14-3-3 proteins acts as chaperones, these interactions may represent a process in which the prey proteins achieve proper protein folding, or OsGF14-c may be responsible for proper subcellular localization of OsPN29982 and OsPN30974. Because all other interactors for OsGF14-c appear to be membrane-associated proteins, OsPN29982 and OsPN30974 are likely to be membrane proteins and may reside at the thylakoid or other cellular membrane structures.

In summary, some of the rice proteins found to interact with OsGF14-c appear to be located at the thylakoid membrane where they participate in photosynthetic processes occurring in the chloroplast; these interactions are consistent with previously reported

localization of 14-3-3 proteins to the chloroplast stroma and the stromal side of thylakoid membranes (Sehnke *et al.*, *Plant Physiol.* **122**(1): 235-242, 2000). Other interactors identified are associated with the plasma or vacuolar membrane. OsGF14-c is, thus, likely to be a membrane component in rice. Because 14-3-3 proteins participate in many types of signaling pathways and are thought to act as molecular chaperones necessary for the assembly, unfolding or transport of proteins through membranes, it is likely that OsGF14-c functions as a molecular glue or stabilizer to regulate the function of the proteins with which it interacts at the thylakoid or other membrane structures. The identification of OsGF14-c as a membrane component represents a novel observation and the first functional characterization of the GF14-c protein in rice. In particular, the proteins identified in this Example as interacting at the thylakoid membrane of chloroplasts represent a novel rice protein complex.

Three interactors were identified in this study for OsDAD1. One is the putative plasma membrane H^+ -ATPase (OsPN23022) that interacts with OsGF14-c. Evidence exists that both OsDAD1 and H^+ -ATPase are integral membrane proteins (Lindholm *et al.*, *Mech. Dev.* **93**(1-2): 169-173, 2000; Ratajczak *et al.*, *Biochim Biophys Acta* **1465**(1-2): 17-36, 2000). H^+ -ATPase translocates protons into intracellular organelles or across the plasma membrane of specialized cells, its activity resulting in acidification of intracellular compartments in eukaryotic cells. The acidic interior of lysosomes has been shown to be necessary for apoptosis under some conditions (Kagedal *et al.*, *Biochem J.* **359**(Pt 2): 335-343, 2001; Bursch, W., *Cell Death Differ.* **8**(6): 569-81, 2001). Thus, the activities of these two enzymes may be necessary for regulation of programmed cell death, and their physical interaction may represent a step in control of this event. Furthermore, 14-3-3 proteins have been implicated in regulation of many cellular processes including apoptosis (van Hemert *et al.*, *Bioessays* **23**(10): 936-946, 2001). It is possible that the interactions of OsPN23022 with GF14-c and with OsDAD1 represent steps in such regulation.

Another novel interactor found for OsDAD1 is the novel rice Na^+ -dependent inorganic phosphate cotransporter. We speculate that the rice phosphate cotransporter is

also a membrane protein based on functional homology with its mammalian homologs, which are localized to neuronal and endocrine vesicles and have a role in glutamate storage (Takamori *et al.*, *Nature* 407(6801): 189-94, 2000). It is likely that glutamate participates in apoptosis regulation in plants as it does in mammals (Bezzi *et al.*, *Nat.*
5 *Neurosci.* 4(7): 702-710, 2001), and that this occurs in rice through the association of the phosphate cotransporter OsPN23053 with OsDAD1.

Finally, OsDAD1 was found to interact with the rice beta-expansin. Expansins are localized to the plasma membrane adjacent to the cell wall, from which they mediate
10 cell wall extension. Since genes regulating cell death are part of the defense response, this interaction may be associated with structural changes in the cell wall in response to cell death.

The interactions here reported represent the first characterization of the DAD1
15 protein homolog in rice. Notably, the fact that OsDAD1 and its interactors appear to be membrane proteins and that one of them, OsPN23022, interacts with OsGF14-c lend further support to the notion that OsGF14-c is a membrane component.

Example VI

20 The rice senescence-associated protein (Os006819-2510) shares 61.4% amino acid sequence similarity with daylily Senescence-Associated Protein 5, a protein encoded by one (DSA5) of six cDNA sequences the levels of which increase during petal senescence. Transcripts of these genes are found predominantly in petals, their
25 expression increase during petal but not leaf senescence, and they are induced by a concentration of abscisic acid (ABA) that causes premature senescence of the petals. Petal senescence is an example of endogenous programmed cell death, or apoptosis, a process in which unwanted cells are eliminated during growth and development. Genes performing a regulatory function in cell death or survival are important to developmental processes. The rice senescence-associated protein Os006819-2510 was chosen as a bait
30 for these interaction studies based on its potential relevance to plant growth and development.

To identify proteins that interacted with the rice senescence-associated protein Os006819-2510, an automated, high-throughput yeast two-hybrid assay technology (provided by Myriad Genetics Inc., Salt Lake City, UT) was employed, as has been described above.

5

Results

The rice senescence-associated protein Os006819-2510 was found to interact with eight rice proteins. Five interactors are known, namely, the rice histone deacetylase HD1 (OsAAK01712), an enzyme involved in regulation of core histone acetylation; the
10 calcium-binding protein calreticulin precursor (OsCRTC), which also interacts with the starch biosynthetic enzyme soluble starch synthase (OsSSS) and with a novel protein (OsPN29950) of unknown function; low temperature-induced protein 5 (OsLIP5); the dehydrin RAB 16B, which is induced by water stress; and rice putative myosin (OsPN23878), an actin motor protein which also interacts with a putative calmodulin-
15 kinase that is associated with a network of proteins involved in cell cycle regulation (see Examples I and II). Three interactors for senescence-associated protein are novel proteins including a putative callose synthase (OsPN23226), an enzyme involved in the biosynthesis of the glucan callose; a protein similar to barley coproporphyrinogen III oxidase, chloroplast precursor, an enzyme of the chlorophyll biosynthetic pathway
20 (OsPN23485); and a protein similar to *Arabidopsis* Gamma Hydroxybutyrate Dehydrogenase.

The interacting proteins of this Example are listed in Table 23, followed by detailed information on each protein and a discussion of the significance of the interactions. The nucleotide and amino acid sequences of the proteins of the Example are
25 provided in Figure 13.

Note that several prey proteins identified are, like the bait protein Os006819-2510, membrane-associated molecules (OsCRTC, OsPN23226, OsLIP5). Several appear to be associated with cell cycle processes in rice (OsPN23878, Os003118-3674,
30 OsCRTC, OsSSS, OsPN23226, OsAAK01712), while others are involved in the plant stress response (OsRAB16B, OsLIP5, OsCRTC). Some of the proteins identified

represent rice proteins previously uncharacterized. Based on the presumed biological function of the prey proteins and on their ability to specifically interact with the bait protein Os006819-2510, Os006819-2510 is speculated to be involved in cell cycle/mitotic processes and in the plant resistance to stress, and may actually represents a link between these processes in rice.

Proteins that participate in cell cycle regulation in rice may be targets for genetic manipulation or for compounds that modify their level or activity, thereby modulating the plant cell cycle. The identification of genes encoding these proteins may allow genetic changes in plant development or growth. Likewise, genes that are involved in conferring plants resistance to stress have important commercial applications, as they could be used to facilitate the generation and yield of crops.

Table 23. Interacting Proteins Identified for Os006819-2510 (Hypothetical Protein 006819-2510, Similar to *Hemerocallis* Senescence-Related Protein 5).

The Myriad names and the TMRI names of the clones of the proteins used as baits and found as preys are given. Nucleotide/protein sequence accession numbers for the proteins of the Example (or related proteins) are shown in parentheses under the protein name. The bait and prey coordinates (Coord) are the amino acids encoded by the bait fragment(s) used in the search and by the interacting prey clone(s), respectively. The source is the library from which each prey clone was retrieved.

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (source)
BAIT PROTEIN :			
Os006819-2510 PN20462	Hypothetical Protein 006819-2510, Similar to Senescence-Related Protein 5 from <i>Hemerocallis</i> Hybrid Cultivar (AAC34855.1; e ⁹⁷)		
INTERACTORS:			
OsAAK01712 PN24059	<i>O. sativa</i> Histone Deacetylase HD1 (AF332875; AAK01712.1)	1-150	90-221 (output trait)
OsCRTC* PN20544	<i>O. sativa</i> Calreticulin Precursor (AB021259; BAA88900)	1-273	283-301 (output trait)
OsLIP5 PN22883	<i>Oryza sativa</i> Low Temperature-Induced Protein 5 (AB011368; BAA24979.1)	1-150	29-60 (input trait)
OsPN23878#	<i>Oryza sativa</i> Putative Myosin (AC090120; AAL31066.1)	1-150	685-888 (output trait)
OsRAB16B PN20554	<i>O. sativa</i> DEHYDRIN RAB 16B (P22911)	1-273	147-164 (output trait)
OsPN23226	Novel Protein PN23226, Callose synthase	1-273	345-432

			(output trait)
OsPN23485	Novel Protein PN23485, Similar to <i>Hordeum vulgare</i> Coproporphyrinogen III Oxidase, chloroplast precursor (Q42840; e ⁻¹⁶⁹)	1-273	90-243 (output trait)
OsPN29037	Novel Protein PN29037	1-150	73-165 (input trait)

* Additional interactions identified for OsCRTC are listed in Table 24

Additional interactions identified for OsPN23878 are listed in Table 25

Table 24

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (source)
BAIT PROTEIN:			
OsCRTC PN20544	Calreticulin Precursor (AB021259; BAA88900)		
INTERACTORS :			
OsPN29950	Novel Protein PN29950	1-150	7-103 2x 138-343 50-343 (output trait)
OsSSS PN19701	Soluble Starch Synthase (AF165890; AAD49850)	250-425	68-270 (input trait) 97-263 (output trait)

5

Table 25

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (source)
PREY PROTEIN:			
OsPN23878	<i>Oryza sativa</i> Putative Myosin (AC090120; AAL31066.1)		
BAIT PROTEIN:			
Os003118-3674 PN20551	Hypothetical Protein 003118-3674 Similar to <i>Lycopersicon esculentum</i> Calmodulin	75-149	824-935 (output trait)

Os006819-2510 is a 276-amino acid protein that includes a cleavable signal peptide (amino acids 1 to 27) and three transmembrane domains (amino acids 48 to 64, 82 to 98, and 233 to 249), as predicted by analysis of its amino acid sequence. The analysis also predicted two endoplasmic reticulum retention motifs, one N-terminal (AFRL) and the other C-terminal (KGGY), and a prokaryotic membrane lipoprotein lipid attachment site beginning with amino acid 57 (Prosites). This site, when functional, is a region of protein processing. Analysis by Pfam also identified a transmembrane superfamily domain, also called a tetraspanin family domain, typically found in a group

of eukaryotic cell surface antigens that are evolutionarily related and include transmembrane domains.

A BLAST analysis against the Genpept database indicated that Os006819-2510 is similar to Senescence-Associated Protein 5 from *Hemerocallis* hybrid cultivar (daylily) (GenBank Accession No. AAC34855.1; 61.4% identity; e^{-97}). In agreement with this result, the protein with the amino acid sequence most similar (63% identity) to that of Os006819-2510 in Myriad's proprietary database is Hypothetical Protein 005991-3479, Similar to *Hemerocallis* Senescence-Associated Protein 5 (Os005991-3479). In an effort to identify the components of the genetic program that leads daylily petals to senescence and cell death ca. 24 hours after the flower opens, the cDNA encoding senescence-associated protein 5 in petals was isolated as one of six cDNAs (designated DSA3, 4, 5, 6, 12 and 15) whose levels increase during petal senescence (Panavas *et al.*, *Plant Mol. Biol.* 40(2): 237-248, 1999). However, no sequence homology was identified in the public database for the DSA5 gene product, which remains as yet unidentified. The levels of DSA mRNAs in leaves was determined to be less than 4% of the maximum detected in petals, with no differences between younger and older leaves, and the DSA genes (except DSA12) are expressed at low levels in daylily roots and (except DSA4) induced by a concentration of abscisic acid that causes premature senescence of the petals.

Two-bait fragments, encoding amino acid 1-273 and 1-150, of Os006819-2510 were used in the yeast two-hybrid screen.

A bait fragment encoding amino acids 1-150 of Os006819-2510 was found to interact with *O. sativa* histone deacetylase HD1 (OsAAK01712). A BLAST analysis of the amino acid sequence of OsAAK01712 indicated that this prey protein is the rice Histone Deacetylase HD1 (GenBank Accession No. AAK01712.1, 100% identity, $e=0.0$). Histone deacetylase (HD) enzymes have been isolated from plants, fungi and animals (reviewed by Lechner *et al.*, *Biochim Biophys Acta* 1296(2): 181-188, 1996). The enzymatic activity of histone deacetylase and that of histone acetyltransferase maintain the enzymatic equilibrium of reversible core histone acetylation. Core histones are a group of highly conserved nuclear proteins in eukaryotic cells; they represent the main

component of chromatin, the DNA-protein complex in which chromosomal DNA is organized. Besides their role in chromatin structural organization, core histones participate in gene regulation, their regulatory function being ascribed to their ability to undergo reversible posttranslational modifications such as acetylation, phosphorylation, glycosylation, ADP-ribosylation, and ubiquitination. Histone deacetylase exists as multiple enzyme forms, and this multiplicity reflects the complex regulation of core histone acetylation. Four nuclear HDs have been identified and characterized from germinating maize embryos (HD1-A, HD1-BI, HD1-BII, and HD2), based on their expression during germination, molecular weight, physiochemical properties and inhibition by various compounds. Based on these data, Lechner *et al.*, *supra*, suggest that HD enzymes have a role in establishing and maintaining histone-protein interactions, and that acetylation may modulate the binding of proteins with anionic domains to certain chromatin areas.

Os006819-2510 was found to interact with *O. sativa* Calreticulin Precursor (OsCRTC). A BLAST analysis of the amino acid sequence of the prey clone OsCRTC indicated that this protein is the rice Calreticulin Precursor (GenBank Accession No. BAA88900/SwissProt #Q9SLY8, 100% identity, $e=0.0$). OsCRTC is a 424-amino acid protein with a cleavable signal peptide (amino acids 1 to 29), a calreticulin family repeat motif (amino acids 218 to 230), and an endoplasmic reticulum targeting sequence (amino acids 421 to 424), as predicted by analysis of the OsCRTC amino acid sequence (see Munro and Pelham, *Cell* 48: 899-907, 1987; Pelham H.R.B., *Trends Biochem. Sci.* 15: 483-486, 1990). In agreement with its designation as a calreticulin precursor, the analysis identified a calreticulin family signature calreticulin family signature (amino acids 31 to 343, $1.3e^{-166}$) (see Michalak *et al.*, *Biochem. J.* 285: 681-692, 1992; Bergeron *et al.*, *Trends Biochem. Sci.* 19: 124-128, 1994; Watanabe *et al.*, *J. Biol. Chem.* 269: 7744-7749, 1994). The analysis also predicted a transmembrane domain (amino acids 7 to 29) and a coiled coil (amino acids 360 to 389). The cDNA encoding the rice calreticulin OsCRTC was first identified by Li and Komatsu, *Eur. J. Biochem.* 267(3): 737-745, 2000 who found this gene to be involved in the regeneration of rice cultured suspension cells. These authors report that the rice calreticulin protein is highly conserved, showing high

homology (70–93%) to other plant calreticulins, but only 50–53% homology to mammalian calreticulins. Calreticulin (CRT) is an endoplasmic reticulum (ER) calcium-binding protein thought to be involved in many functions in eukaryotic cells, including Ca^{2+} signaling, regulation of intracellular Ca^{2+} storage and store-operated Ca^{2+} fluxes through the plasma membrane, modulation of endoplasmic reticulum Ca^{2+} -ATPase function, chaperone activity to promote protein folding, control of cell adhesion, gene expression, and apoptosis (reviewed by Michalak *et al.*, *Biochem. Cell. Biol.* **76**(5): 779–785, 1998 and by Persson *et al.*, *Plant Physiol.* **126**(3): 1092–1104, 2001). In plants, CRT has been localized to the endoplasmic reticulum, Golgi, plasmodesmata, and plasma membrane (Borisjuk *et al.*, *Planta* **206**(4): 504–14, 1998; Hassan *et al.*, *Biochem. Biophys. Res. Commun.* **211**(1): 54–49, 1995; Baluska *et al.*, *Plant Physiol.* **126**(1): 39–46, 2001), and it has been shown to affect cellular calcium homeostasis, as reported by Persson *et al.*, *supra*. This study shows that induction of calreticulin expression in transgenic tobacco and *Arabidopsis* plants enhances the ATP-dependent Ca^{2+} accumulation of the endoplasmic reticulum, and that this CRT-mediated alteration of the ER Ca^{2+} pool regulates ER-derived Ca^{2+} signals. These results demonstrate that CRT plays a key role as a regulator of calcium storage in the endoplasmic ER, and that the ER, in addition to the vacuole, is an important Ca^{2+} store in plant cells. A role for the *Arabidopsis* calreticulin homolog in anther maturation or dehiscence has also been proposed (Nelson *et al.*, *Plant Physiol.* **114**(1): 29–37, 1997) based on localization of this protein in anthers which are degenerating at the time of maximum CRT expression. Furthermore, the tobacco homolog of mammalian CRTC participates in protein-protein interactions in a stress- and ATP-dependent fashion Denecke *et al.*, *Plant Cell* **7**(4): 391–406, 1995). This notion supports the use of the yeast two-hybrid technology to identify proteins that interact with OsCRTC.

OsCRTC was also used as bait and found to interact with rice Soluble Starch Synthase (OsSSS) (see Table 24) and Novel Protein PN29950 (OsPN29950). OsSSS is the rice homolog of soluble starch synthase (SSS), one of the three enzymes involved in starch biosynthesis in plants. Starch is the major component of yield in the world's main crop plants and one of the most important products synthesized by plants that is used in

industrial processes. It consists of two kinds of glucose polymers: highly branched amylopectin and relatively unbranched amylose. Starch synthase contributes to the synthesis of amylopectin. The enzyme utilizes the glucosyl donor ADPGlc to add glucosyl units to the nonreducing end of a glucan chain through $\alpha(1 \rightarrow 4)$ linkages, thus
5 elongating the linear chains (reviewed by Cao *et al.*, *Arch Biochem Biophys.* 373(1): 135-46, 2000; Kossman and Lloyd, *Crit. Rev. Biochem. Mol. Biol.* 35(3): 141-196, 2000).

Distinct classes of isoforms of starch synthase were defined on the basis of similarity in amino acid sequence, molecular mass, and antigenic properties. Plant organs vary greatly in the classes they possess and in the relative contribution of the classes to soluble starch
10 synthase activity (Smith *et al.*, *Ann. Rev. Plant Biol.* 48(1): 67, 1997 cited in Cao *et al.*, *supra*). OsPN29950 is a protein of unknown function determined by BLAST analysis to be similar to putative protein from *Arabidopsis thaliana* (GenBank Accession No. NP_199037.1, 32% identity, $2e^{-29}$).

15 Os006819-2510 was found to interact with low temperature-induced protein 5 (OsLIP5). OsLIP5 is a 276-amino acid protein with a cleavable signal peptide (amino acids 1 to 27) and three putative transmembrane regions (amino acids 48 to 64, 82 to 98, and 233 to 249). A BLAST analysis of the amino acid sequence of this prey clone
20 determined that it is the rice LIP5 protein (GenBank Accession No. BAA24979.1, 100% identity, $8e^{-052}$). The rice LIP5 protein is a direct submission to the public database and is not described in the literature. In yeast, LIP5 is involved in lipoic acid metabolism (Sulo and Martin, *J. Biol. Chem.* 268(23): 17634-17639, 1993). The BLAST analysis shows that the rice LIP5-like protein OsLIP5 is also similar to rice WSI724 (Accession #T07613, 98% identity, $3e^{-051}$), a protein encoded by one of nine cDNAs induced by
25 short-term water stress and thought to be responsible for acquired resistance to chilling in a chilling-sensitive variety of rice (Takahashi *et al.*, *Plant Mol. Biol.* 26(1): 339-352, 1994). Among the proteins encoded by these cDNAs, which were found to be differentially expressed following water stress, expression of the WSI724 protein remained relatively fixed. A BLAST analysis comparing the nucleotide sequence of
30 OsLIP5 against TMRI's GeneChip[®] Rice Genome Array sequence database identified

probeset OS000070_r_at ($e=4e-75$) as the closest match. Gene expression experiments indicated that this gene is down-regulated by the herbicide BL2.

Os006819-2510 was also found to interact with *Oryza sativa* putative myosin
5 (OsPN23878). A BLAST analysis of the amino acid sequence of OsPN23878 indicated
that this prey protein is the rice putative myosin (GenBank Accession No. AAL31066.1,
99% identity, $e=0.0$). OsPN23878 is also similar to Myosin VIII, ZMM3 - maize
(fragment) from *Z. mays* (GenBank Accession No. A59311, 89% identity, $e=0.0$).
Myosins are discussed in Example I. Based on current knowledge of plant myosins, the
10 myosin VIII prey protein OsPN23878 may be a cytoskeletal component that participates
in events relating to cytokinesis.

The prey protein OsPN23878 also interacts with hypothetical protein 003118-
3674, which is similar to *Lycopersicon esculentum* Calmodulin (Os003118-3674) (see
Table 25). Os003118-3674 is a 148-amino acid protein with two EF-hand calcium-
15 binding domains (amino acids 22 to 34 and 93 to 105). In agreement with the
observation that Os003118-3674 includes EF-hand calcium-binding domains, a BLAST
analysis of the Genpept database indicated that this protein shares 72% identity with *A.*
thaliana putative calmodulin (GenBank Accession No. NP_1764705, e^{-57}), although the
top hit in this search is *A. thaliana* putative serine/threonine kinase (GenBank Accession
20 No. NP_172695.1, 76% identity, $7e^{-60}$). Therefore, the possibility that this calmodulin-
like protein possesses kinase activity is worth consideration.

A BLAST analysis comparing the nucleotide sequence of OsPN23878 against
TMRI's GeneChip® Rice Genome Array sequence database identified probeset
OS002190_I_at ($e=-165$) as the closest match. Our gene expression experiments indicate
25 that this gene is not specifically induced under a range of given conditions.

Additionally, Os006819-2510 was found to interact with OsRAB16B
(OsRAB16B), a 164-amino acid protein that has a possible cleavage site between amino
acids 51 and 52, although it does not appear to have a cleavable signal peptide. Analysis
30 of its amino acid sequence predicted ($2.6e^{-81}$) this protein to be a member of a group of
plant proteins called dehydrins, which are induced in plants by water stress (see (Close *et*

al., *Plant Mol. Biol.* **13**: 95-108, 1989; Robertson and Chandler, *Plant Mol. Biol.* **19**:
 1031-1044, 1992; Dure *et al.*, *Plant Mol. Biol.* **12**: 475-486, 1989). Dehydrins include
 the basic, glycine-rich RAB (responsive to abscisic acid) proteins. In agreement with this
 notion, the analysis indicated that OsRAB16B is a basic, glycine-rich protein. A BLAST
 5 analysis against the public database revealed that OsRAB16B is the rice DEHYDRIN
 RAB 16B (GenBank Accession No. P22911, 100% identity, $4e^{-95}$). The cDNA encoding
 this protein was isolated by (Yamaguchi-Shinozaki *et al.*, *Plant Mol. Biol.* **14**(1): 29-39,
 1990) as one of four rice RAB genes that are differentially expressed in rice tissues. In
 agreement with the notion that OsRAB16B is a rice RAB protein, a BLAST analysis
 10 against Myriad's proprietary database indicated that OsRAB16B shares 57% identity
 with OsRAB25. While expression data for OsRAB16B are not available, the rice
 RAB16B promoter contains two abscisic acid (ABA)-responsive elements required for
 ABA induction (Ono *et al.*, *Plant Physiol.* **112**(2): 483-491, 1996). Among other rice
 RAB proteins, the RAB16A gene has been linked to salt stress (Saijo *et al.*, *Plant Cell*
 15 *Physiol.* **42**(11): 1228-1233, 2001), and the activity of the RAB16A promoter is also
 induced by ABA and by osmotic stresses in various tissues of vegetative and floral
 organs (Ono *et al.*, *supra*). Another rice RAB protein, RAB21, is induced in rice
 embryos, leaves, roots and callus-derived suspension cells treated with NaCl and/or ABA
 (Mundy and Chua, *EMBO J.* **7**(8): 2279-2286, 1988). Based on these data, it is likely
 20 that the OsRAB16B prey protein has a role in the stress response.

Os006819-2510 was found to interact with protein PN23226 (OsPN23226).
 A BLAST analysis against the public database indicated that OsPN23226 is similar to
 putative glucan synthase (GenBank Accession No. NP_563743.1, 78% identity, $e=0.0$)
 25 and to callose synthase 1 catalytic subunit (GenBank Accession No. NP_563743.1, 78%
 identity, $e=0.0$) from *A. thaliana*. Callose synthase (CalS) from higher plants is a
 multisubunit membrane-associated enzyme involved in callose synthesis (reviewed in
 Hong *et al.*, *Plant Cell* **13**(4): 755-768, 2001). Callose is a linear 1,3- β -glucan with some
 1,6- branches and differs from cellulose, the major component of the plant cell wall.
 30 Callose is synthesized on the forming cell plate and several other locations in the plant,
 and its deposition at the cell plate precedes the synthesis of cellulose. Callose synthesis

can also be induced by wounding, pathogen infection, and physiological stress. The activity of callose synthase is highly regulated during plant development and may be affected by various biotic and abiotic factors. CalS, like cellulose synthase, is a large transmembrane protein. Its structure includes a large hydrophilic loop that is relatively conserved among the CalS isoforms, a less conserved, long N-terminal segment, and a short C-terminal segment, all located on the cytoplasmic side. The central loop is thought to act as a receptacle to hold other proteins that are essential for CalS catalytic activity (see below); the N-terminal segment may contain subdomains for interaction with proteins that regulate 1,3- β -glucan synthase activity.

The cDNA encoding the callose synthase (CalS1) catalytic subunit from *Arabidopsis* was identified by Hong *et al.*, *supra*), who demonstrated that higher plants encode multiple forms of CalS enzymes and that the *Arabidopsis* CalS1 is a cell plate-specific isoform. In addition, these authors used yeast two-hybrid and *in vitro* experiments to show that CalS1 interacts with two other cell plate-specific proteins, phragmoplastin and a UDP-glucose transferase, and suggest that it may form a large complex with these and other proteins to facilitate callose deposition on the cell plate. Moreover, the plasma membrane CalS is strictly Ca^{2+} -dependent, and Ca^{2+} plays a key role in cell plate formation and may activate the cell plate-specific CalS1. The prey protein OsPN23226 is likely a rice callose synthase homolog that may function similarly to the *Arabidopsis* CalS1 catalytic subunit.

In addition to the cell plate, callose is synthesized in a variety of specialized tissues and in response to mechanical and physiological stresses. Multiple CalS isozymes are thought to be required in higher plants to catalyze callose synthesis in different locations and in response to different physiological and developmental signals (Hong *et al.*, *supra*).

Os006819-2510 was also found to interact with protein PN23485, which is similar to *Hordeum vulgare* coproporphyrinogen III oxidase, chloroplast precursor (OsPN23485). A BLAST analysis of the amino acid sequence of OsPN23485 determined that this protein is similar to barley (*H. vulgare*) Coproporphyrinogen III Oxidase, Chloroplast Precursor (coprogen oxidase) (GenBank Accession No. Q42840,

89.3% identity, e⁻¹⁶⁹). Coproporphyrinogen III oxidase (CPO) catalyzes a step in the pathway from 5-amino-levulinate to protoporphyrin IX, a common reaction in the biosynthesis of heme in animals and chlorophyll in photosynthetic organisms. The N-terminal sequences of plant CPOs are characteristic of plastid transit peptides. CPO is
5 exclusively located in the stroma of plastids, and *in vitro* transcribed and translated CPO is imported into the stroma of pea plastids and truncated by a stromal endopeptidase (reviewed by Ishikawa *et al.*, *Plant J.* 27(2): 89-99, 2001). Plant cDNA sequences encoding CPO were obtained from soybean, tobacco and barley (Kruse *et al.*, *Planta* 196(4): 796-803, 1995). They found that the plant coprogen oxidase mRNA was
10 expressed to different extents in various tissues, with maximum amounts in developing cells and drastically decreased amounts in completely differentiated cells, suggesting differing requirements for tetrapyrroles in different organs. Based on these results, these authors propose that enzymes involved in tetrapyrrole (porphyrin) synthesis are regulated developmentally rather than by light, and that regulation of these enzymes guarantees a
15 constant flux of metabolic intermediates and help avoid photodynamic damage by accumulating porphyrins. Inhibition of the pathway for chlorophyll synthesis causes lesion formation such as that found in the pale green and lesion-formation phenotype of *lin2* plants. Ishikawa *et al.*, *supra* found that a deficiency of coproporphyrinogen III oxidase causes lesion formation in these *Arabidopsis* mutants. Furthermore, based on the
20 observation that transgenic tobacco plants with reduced CPO activity accumulate photosensitizing tetrapyrrole intermediates and exhibit antioxidative responses and necrotic leaf lesions, these authors suggest that CPO inhibition causes lesion formation leading to induction of a set of defense responses that resemble the HR observed after pathogen attack. These lesions are the equivalent of diseases known as porphyrias in
25 humans. If accumulated, coproporphyrin(ogen), as a photosensitizer, induces damage through generation of reactive oxidative species, which play a key role in the initiation of cell death and lesion formation both in the HR and in certain lesion mimic mutants. They suggest that in *lin2* mutants, the generation of an oxidative burst triggered by coproporphyrin accumulation leads to cell death.

30

Os006819-2510 was found to interact with protein PN29037 (OsPN29037). A BLAST analysis of the amino acid sequence of OsPN29037 indicated that this prey protein is similar to Gamma Hydroxybutyrate Dehydrogenase from *A. thaliana* (GenBank Accession No. AAK94781.1, 80.7%, identity, e⁻¹²⁷). This enzyme oxidizes gamma-hydroxybutyrate. As a minor brain metabolite directly or indirectly involved in scavenging oxygen-derived free radicals in animals, gamma-hydroxybutyrate demonstrates similarities with melatonin (Cash C.D., *Med Hypotheses* 47(6): 455-459, 1996).

10 Summary

Thus, the senescence-associated protein Os006819-2510 interacts with several proteins that have possible roles in cell cycle processes. One of these is OsPN23878, a protein annotated in the public domain as the rice putative myosin. Myosins are cytoskeletal proteins that function as molecular motors in ATP-dependent interactions with actin filaments in various cellular events. Based on the similarity of the prey protein to a class VIII myosin and on the reported role of plant myosin VIII in maturation of the cell plate and in organization of the actin cytoskeleton at cytokinesis, we speculate that the myosin OsPN23878 is a cytoskeletal component that participates in events occurring at cytokinesis in rice. The association of the myosin OsPN23878 with senescence-associated protein may be a step in cell-cycle-dependent events involving cytoskeleton organization and senescence. Specific expression of the gene encoding OsPN23878 in panicle (our gene expression experiments) is consistent with an interaction between this protein and Os006819-2510, and with a role for the latter in flower senescence, as suggested for the gene encoding the daylily homolog of this protein (Panavas *et al.*, *Plant Mol. Biol.* 40(2): 237-248, 1999). Localization of senescence-associated protein to the ER suggests that some of the events in which OsPN23878 functions could be associated with plasmodesmata function.

Note that the myosin protein OsPN23878 also interacts with a novel calmodulin-kinase-like protein Os003118-3674 (see Table 25), and that the latter interacts with a myosin heavy chain (OsAAK98715) found to interact with rice cyclin OsCYCOS2 and

presumed to be involved in cytoskeleton organization during mitotic events (see Example II). The interactions of myosins with a calcium-binding calmodulin-like protein are consistent with published evidence of regulation of myosin function by calcium (Yokota *et al.*, *Plant Physiol.* **119**(1): 231-240, 1999, reviewed in Reddy A.S., *Int. Rev. Cytol.* **204**: 97-178, 2001). The possibility that Os003118-3674 possesses kinase activity raises the probability that these interactions propagate a cell-cycle-related signaling event. The calmodulin-like protein Os003118-3674 thus provides a link between the senescence-associated protein and interacting partners of this Example and the cell cycle network.

Another interactor with a possible role in cell cycle regulation is the rice histone deacetylase OsAAK01712. This enzyme includes a transmembrane domain and is involved in regulation of core histones acetylation. The acetylation/deacetylation of histones, the main protein component of chromatin, is connected to replication during the cell cycle in plants, as is in other eukaryotes (Jasenčakova *et al.*, *Chromosoma* **110**(2): 83-92, 2001). Thus, the Os006819-2510-OsAAK01712 interaction likely participates in mitotic events involving chromatin organization.

Another novel interactor found for senescence-associated protein is OsPN23485, similar to coproporphyrinogen III oxidase, chloroplast precursor, an enzyme of the pathway leading to the biosynthesis of chlorophyll in plants. The observation that the lesion formation in the lin2 mutant *Arabidopsis* plants is the result of loss-of-function of CPO (Ishikawa *et al.*, *Plant J.* **27**(2): 89-99, 2001) links the gene encoding CPO to regulation of cell death pathways. Moreover, plant CPO enzymes are regulated developmentally and by light (reviewed by Ishikawa *et al.*, *supra*). Based on these reports, the interaction of rice CPO (OsPN23485) with senescence-associated protein may participate in regulation of programmed cell death in a development-dependent manner in rice.

The senescence-associated protein Os006819-2510, which is presumed to be a transmembrane protein based on analysis of its amino acid sequence, interacts with the rice calreticulin OsCRTC which, like other plant calreticulins, is likely an ER

25 To further corroborate the notion that the rice senescence-associated protein
Os006819-2510 is a membrane-associated protein, a novel interactor identified for this
protein is a putative callose synthase catalytic subunit (OsPN23226), another
transmembrane enzyme involved in glucan synthesis. Plasma membrane proteins
participate in a variety of interactions with the cell wall, including synthesis and assembly
30 of cell wall polymers (Biochemistry and Molecular Biology of Plants, Buchanan,
Gruissem and Jones (eds.), John Wiley & Sons, New York, NY 2002, p. 13). The prey

protein OsPN23226 likely functions as its *Arabidopsis* homolog, a plasma membrane enzyme that utilizes UDP-glucose as substrate to synthesize callose for deposition in the cell wall. The interactions of senescence-associated protein with the rice putative callose synthase OsPN23226 and with the calreticulin OsCRTC, and the interaction between
5 OsCRTC and the soluble starch synthase OsSSS all involve membrane-associated proteins. While there is no evidence that such interactions occur at the same time, they may be associated with the traffic that sorts, distributes and targets membrane proteins and other molecules between compartments of the endomembrane system (Biochemistry and Molecular Biology of Plants, Buchanan, Gruissem and Jones (eds.), John Wiley &
10 Sons, New York, NY 2002, p. 14) during the different stages of the cell cycle/development and in response to different physiological and developmental signals. Moreover, the interactions identified in this Example link the senescence-associated bait protein to glucan synthesis, a process that is vital to the plant normal growth. For example, the formation of a functional callose synthase 1 catalytic subunit (CalS1)
15 complex is vital to cell plate formation. Functional characterization of the various components of the CalS1 complex and CalS-associated proteins has been proposed as a means to reveal how the activity of this enzyme is regulated during cell plate formation and to clarify callose synthesis and deposition in plants (Hong et al., *Plant Cell* 13(4): 755-768, 2001). The interaction identified here between senescence-associated protein
20 and the novel putative callose synthase catalytic subunit (OsPN23226) provides new insight into this process in rice.

Other interactors identified for senescence-associated protein link this protein to the plant stress response. OsRAB16B is a member of the RAB family of proteins known
25 to be induced by water stress and treatment with the plant hormone abscisic acid. ABA levels increase during seed development in many plant species, stimulating production of seed storage proteins and preventing premature germination; ABA is also induced by water stress and is thought to regulate stomatal transpiration (Raven, Eivert and Eichhorn, p. 684). Based on functional homology with other RAB proteins and on the presence of
30 the ABA-responsive elements in the OsRAB16B promoter, we presume that OsRAB16B has a role in the response to abiotic stress in rice and that its function may be regulated by

Ca²⁺. Another interactor correlated with stress is low temperature-induced protein 5 (OsLIP5), which in yeast is involved in lipoic acid metabolism. Lipoic acid in animals has been shown to help minimize the effects of systemic stress (Kelly G.S., *Altern. Med. Rev.* 4(4): 249-265, 1999) and to provide animal cells with significant protection against the cytotoxic effects of repin, a sesquiterpene lactone isolated from Russian knapweed (Robles *et al.*, *J. Neurosci. Res.* 47(1): 90-97, 1997). The high similarity (98%) of the rice LIP5-like protein to rice WSI724, a protein encoded by a gene induced by water stress and linked to resistance to chilling in rice, points to similar roles for the OsLIP5 prey protein. Gene expression experiments indicate that the gene encoding OsLIP5 is down-regulated upon treatment with the herbicide BL2. This finding suggests a role for OsLIP5 in the response to abiotic stress. While the specific function of the interactions between Os006819-2510 and the prey proteins OsRAB16B and OsLIP5 is not obvious, these interactions may participate in biological processes related to flower senescence and response to water stress and chilling.

In addition, the rice calreticulin OsCRTC discussed above may also have a role in the stress response. This hypothesis is based on functional homology with the tobacco CRT protein studied by Denecke *et al.* (*Plant Cell* 7(4): 391-406, 1995) and found to participate in protein-protein interactions in a stress-dependent fashion.

In summary, among the interactors identified for the rice senescence-associated protein Os006819-2510 are several membrane-associated proteins, which supports the notion that the rice Os006819-2510 is a transmembrane protein. Among the interactors identified are proteins involved in cell cycle processes/mitosis and proteins with functions in the plant stress response. Some are newly characterized rice proteins. The interactions identified for rice senescence-associated protein with proteins involved in cell cycle/development and in resistance to stress suggests an overlapping of roles for the bait protein. Indeed, Os006819-2510 may constitute a link between stress tolerance and processes for cell division in rice.

Example VII

OsSGT1 is a 367-amino acid protein that includes a tetratricopeptide repeat domain, two variable regions, the CS motif present in metazoan CHORD and SGT1 proteins, and the SGS motif. In yeast, Sgt1 is required for cell-cycle signaling. In yeast, SGT1 associates with the kinetochore complex and the SCF-type E3 ubiquitin ligase by interacting with SKP1. COP9 signalosome interacts with SCF E3 ubiquitin ligases. By its interaction with SCF complexes, SGT1 exerts its essential activity in degrading of SIC1 and CLN1. Thus, one possible role of SGT1 could be to target proteins for degradation by the 26S proteasome via specific SCF complexes or the SGT1 complex may participate in the modification of protein activity or may have a dual role for activation and degradation of the target via ubiquitylation. *A. thaliana* has two SGT1 homologs. At nonpermissive temperatures AtSGT1a and AtSGT1b can complement G1 and G2 arrest in temperature sensitive *sgt1* yeast mutants. However, SGT1b interacts with RAR1 which is required for RPP5 regulated disease resistance to downy mildew. In this scenario, target proteins involved in disease resistance may be targeted for protein degradation by the SGT1 pathway. Barley encodes a SGT1 homolog that also interacts with barley RAR1, which is implicated in disease resistance in barley to downy mildew. (Austin *et al.*, *Science* **295**(5562): 2077-2080, 2002; Azevedo *et al.*, *Science* **295**(5562): 2073-2076, 2002). A BLAST analysis comparing the nucleotide sequence of OsSGT1 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS016424.1 (98%) as the closest match. Gene expression experiments indicated that this gene is up-regulated by the blast infection.

The rice SGT1 protein shares 74 and 75% amino acid sequence similarity with two *Arabidopsis thaliana* SGT1 homologs and 45% amino acid sequence similarity with *Saccharomyces cerevisiae* SGT1. In yeast, SGT1 is required for cell-cycle progression at the G1/S-phase and G2/M-phase transitions. In *A. thaliana*, SGT1b interacts with Rar1 and mediates disease resistance. Thus, in plants, SGT1 likely controls processes that are fundamental to disease resistance and development. The rice OsSGT1 protein was chosen as a bait for these interaction studies based on its potential relevance to disease resistance and development. One bait fragment encoding amino acid 200-368 of OsSGT1 was used in the yeast two-hybrid screen, as described above.

Results

The OsSGT1 was found to interact with ten rice proteins. Three interactors have been previously described, namely OsSGT1, a Ras GTPase (gi|730510), and elicitor responsive protein (gi|11358958). The remaining seven interactors are novel proteins with identifiable protein domains, or are similar to other proteins. These are an L-aspartase-like protein, an RNA binding domain protein, an auxin induced-like protein, an archain delta COP-like protein, a fibrillin-like protein, a HSP70-like protein, and a proline rich protein. The elicitor responsive protein was also used as a bait and interacted with 12 novel proteins with identifiable protein domains, with similarity to known proteins or that are unidentifiable by sequence similarity. These were an NAD(P) binding domain protein, a gamma adaptin-like protein, a pectinesterase-like protein, a receptor like kinase protein kinase like protein, a pyruvate orthophosphate dikinase like protein, an Isp-4 like protein, a xanthine dehydrogenase like protein, a ubiquitin specific protease like protein and 4 unknown proteins.

The interacting proteins of this Example are listed in Table 26, followed by detailed information on each protein and a discussion of the significance of the interactions. The nucleotide and amino acid sequences of the proteins of the Example are provided in Figure 14. Based on the biological function of SGT1, it is possible that the interacting proteins are also involved in cell cycle/mitotic processes and/or in the plant resistance to stress. Likewise, the interactors with the elicitor responsive protein may also be involved in plant resistance to stress. Proteins that participate in cell cycle regulation in rice may be targets for genetic manipulation or for compounds that modify their level or activity, thereby modulating the plant cell cycle. The identification of genes encoding these proteins may allow genetic manipulation of crops or application of compounds to effect agronomically desirable changes in plant development or growth. Likewise, genes that are involved in conferring plants resistance to stress have important commercial applications, as they could be used to facilitate the generation and yield of stress-resistant crops.

Table 26. Interacting Proteins Identified for Os006819-2510 (Hypothetical Protein 006819-2510, Similar to *Hemerocallis* Senescence-Related Protein 5).

The Myriad names and the TMRI names of the clones of the proteins used as baits and found as preys are given. Nucleotide/protein sequence accession numbers for the proteins of the Example (or related proteins) are shown in parentheses under the protein name. The bait and prey coordinates (Coord) are the amino acids encoded by the bait fragment(s) used in the search and by the interacting prey clone(s), respectively. The source is the library from which each prey clone was retrieved.

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (source)
BAIT PROTEIN :			
PN20285	OsSGT1 (gi 6581058)		
INTERACTORS:			
PN24060	L-aspartase-like protein-like	200-368	176-315 (output trait)
PN20696* (OsERP)	Elicitor responsive protein (gi 11358958)	200-368	54-144 (input trait)
PN23914	RNA binding domain protein	200-368	1-263 x 3 (output trait)
PN23221#	Proline rich protein	200-368	182-366 x 2 (output trait) 207-344 (input trait) 134-254 (output trait)
PN20285	OsSGT1 (gi 6581058)	200-368	9-227 (output trait)
PN24061	Auxin induced protein-like	200-368	34-236 (output trait)
PN24063	RAS GTPase (gi 730510)	200-368	63-202 (output trait)
PN23949	HSP70-like	200-368	244-418 (output trait)
PN29042	Fibrillin-like		

* Additional interactions identified for elicitor responsive protein are shown in Table 27

Additional interactions identified for PN23221 are shown in Table 28

Table 27

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (source)
BAIT PROTEIN :			
PN20696 (OsERP)	Elicitor responsive protein (gi 11358958)		
INTERACTORS :			
PN29984	Novel Protein PN29950	50-145	1-38 5-41

			(input trait)
PN30844	Novel protein PN30844	50-145	1-64 (input trait)
PN30868	NAD(P) binding domain protein	50-145	167-336 (input trait)
PN24292	Gamma adaptin-like	23-120	737-918 (output)
PN29983	Novel protein PN29983	50-145	1-131 (input trait)
PN30845	Pectinesterase-like	50-145	1-64 (input trait)
PN31085	Receptor-like protein kinase-like	23-120	378-553 (output trait)
PN20674	Pyruvate orthophosphate dikinase-like	50-145	64-263 71-298 (input trait)
PN30870	Isp-4 like	50-145	1-446 (input trait)
PN29997	Xanthine dehydrogenase-like	23-120	737/918 (output trait)
PN30843	Ubiquitin specific protease-like	50-145	164-221 (input trait)
PN30857	Novel protein PN30857	50-145	1-148 (input trait)

Table 28

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (source)
PREY PROTEIN:			
PN23221	Proline rich protein		
BAIT PROTEIN:			
PN20621	Shaggy kinase (gi13677093)	120-435	175-311 (output trait)
PN20115	Ring zinc finger protein	5-140	84-302 191-324 (output trait)

Yeast Two-Hybrid using OsSGT1 as Bait

The bait fragment encoding amino acid 200-368 of OsSGT1 was found to interact with L-aspartase-like protein PN24060. A BLAST analysis of the amino acid sequence of PN24060 indicated that this prey protein has 36.5% similarity to *A. thaliana* L-aspartase (gi|18394135). The enzyme L-aspartate ammonia-lyase (aspartase) catalyzes the reversible deamination of the amino acid L-aspartic acid, using a carbanion mechanism to produce fumaric acid and ammonium ion. While the catalytic activity of this enzyme has been known for nearly 100 years, a number of recent studies have revealed some interesting and unexpected new properties of this reasonably well-characterized enzyme. The non-linear kinetics that are seen under certain conditions have been shown to be caused by the presence of a separate regulatory site. The substrate, aspartic acid, can also play the role of an activator, binding at this site along with a required divalent metal ion. So it is possible that PN24060 catalyses a reaction that pertains to protein modification and the modification may be important for disease resistance or cell cycling.

The bait fragment encoding amino acid 200-368 of OsSGT1 was also found to interact with elicitor responsive protein, PN20696. A BLAST analysis of the amino acid sequence of the prey clone PN20696 indicated that this protein is the rice elicitor responsive protein (gi|11358958; OsERP). OsERP is a 144-amino acid protein that, according to Genbank, is expressed by rice culture cells in the presence of the rice blast fungal elicitor. Thus, OsERP may have a role in disease responses in rice.

OsERP was also used as bait and found to interact with 12 other proteins (see Table 27). These prey are described in this Example below.

An *A. thaliana* homologue to OsERP was identified by BLAST. At1g63220 shares 75% amino acid similarity with OsERP. To see if Arabidopsis homologues of OsERP have roles in disease resistance, *Arabidopsis thaliana* with T-DNA insertions in At1g63220 (line SAIL_320_D02) was identified from a random insertion seed library. DNA regions surrounding the insertions were sequenced and revealed that the T-DNAs were located within exon 5 of At1g63220. Plants were backcrossed and plants homozygous for the T-DNA insertion were identified by PCR. Homozygous mutants and wild type plants were challenged with *Pseudomonas syringae* pv. *maculicola* ES4326 and

plants were assayed for amount of *P. syringae* bacteria accumulation 3 days post inoculation (Glazebrook *et al.*, *Genetics* 143(2): 973-982, 1996) These experiments were repeated twice on at least six plants. Data are reported as means and standard deviations of the log of colony forming units per leaf cm². By three days after inoculation, the mutant plants accumulated more than 10 times as much bacteria as wild type plants (wt = 3.94 log cfu/leaf disk std. 0.57, *atlg63220* = 5.34 std. 0.63). Hence, *Atlg63220* contributes to disease resistance in *A. thaliana*. It is possible that the *Atlg63220* mutation inhibits defense responses that are dependent upon SGT1 interactions.

In addition, the bait fragment encoding amino acid 200-368 of OsSGT1 was found to interact with RNA-binding domain protein, PN23914. PN23914 is a 164-amino acid protein. A BLAST analysis of the amino acid sequence of this prey shows it has 35.9% sequence identity to tFZR1 from *Oncorhynchus mykiss* (gi|2982698). tFZR1 is an orphan nuclear receptor family member, tFZR1, which has a FTZ-F1 box. The amino acid sequences of the zinc finger domain and the FTZ-F1 box has 92.8% and 100% identity, respectively, with those of zebrafish FTZ-F1. On the other hand, the overall homology between tFZR1 and zebrafish FTZ-F1 is low (33.0%). The results indicate that tFZR1 is a new member of fushitarazu factor 1 (FTZ-F1) subfamily. It is possible that PN23914 shares functionality through the zinc finger domain.

In addition, bait fragment encoding amino acid 200-368 of OsSGT1 was found to interact with proline rich protein, PN23221. A BLAST analysis of the amino acid sequence of PN23221 indicated that this prey protein is 40.3% similar to a rice repetitive proline rich protein (gi|18478606). Proline rich proteins may mediate interaction among proteins (Zhao *et al.*, *EMBO J.* 20(9): 2315-2325, 2001). Note that proline rich protein PN23221 also interacts with shaggy kinase PN20621 and ring zinc finger protein-like PN20115 (see Table 28). Thus, the proline rich protein PN23221 may serve to bring these proteins together with OsSGT1.

The bait fragment encoding amino acid 200-368 of OsSGT1 was also found to interact with OsSGT1. In other words, OsSGT1 interacts with itself. Although the bait for OsSGT1 included amino acids 200-368, the prey included amino acids 9-227.

Although OsSGT1 may be a self-regulator through aggregation, these bait and prey domains may reflect natural protein folding of a single native OsSGT1 protein.

5 Additionally, the bait fragment encoding amino acid 200-368 of OsSGT1 was found to interact with an auxin-induced protein like protein, PN24061. A BLAST analysis against the public database indicated that PN24061 is 63.5% similar to a rice putative IAA1 protein (gi|17154533). Indole acetic acid is a plant growth hormone and is classified as an auxin. IAA is associated with a variety of physiological processes, including apical dominance, tropisms, shoot elongation, induction of cambial cell
10 division and root initiation. Thus, genes that are induced by IAA likely produce proteins that are responding developmental changes. This associated goes hand in hand with regulation of cell division by interaction with SGT1.

 The bait fragment encoding amino acid 200-368 of OsSGT1 was also found to interact with Ras GTPase, PN24063. A BLAST analysis of the amino acid sequence of
15 PN24063 determined that this protein is ras-related GTP binding protein possessing GTPase activity (gi|730510). This protein has four conserved regions involved in GTP binding and hydrolysis which are characteristic in the ras and ras-related small GTP-binding protein genes. In addition, two consecutive cysteine residues near the carboxyl-terminal end required for membrane anchoring are also present. This protein synthesized
20 in *Escherichia coli* possessed GTPase activity (*i.e.*, hydrolysis of GTP to GDP) (Kidou *et al.*, *FEBS Lett.* 332(3): 282-286, 1993). Ras GTPases are likely involved in signaling processes for development. ORFX from tomato that is expressed early in floral development, controls carpel cell number, and has a sequence suggesting structural similarity to the human oncogene c-H-ras p21 (fw2.2: a quantitative trait locus key to the
25 evolution of tomato fruit size. (Frary *et al.*, *Science* 289(5476): 85-88, 2000). The Rho family of GTPases are also involved in control of cell morphology, and are also thought to mediate signals from cell membrane receptors (Winge *et al.*, *Plant Mol. Biol.* 35(4): 483-495, 1997).

30 An *A. thaliana* homologue to PN24063 was identified by BLAST. At1g02130 shares 90% amino acid similarity with PN24063. To see if Arabidopsis homologues of

PN24063 have roles in disease resistance *Arabidopsis thaliana* with T-DNA insertions in *At1g02130* (line SAIL_680_D03) was identified from a random insertion seed library. DNA regions surrounding the insertions were sequenced and revealed that the T-DNAs were located within the promoter of *At1g02130*. Plants were backcrossed and plants
5 homozygous for the T-DNA insertion were identified by PCR. Homozygous mutants and wild type plants were challenged with *Pseudomonas syringae* pv. *maculicola* ES4326 and plants were assayed for amount of *P. syringae* bacteria accumulation 3 days post inoculation (Glazebrook *et al.*, *supra*). These experiments were repeated twice on at least six plants. Data are reported as means and standard deviations of the log of colony
10 forming units per leaf cm². By three days after inoculation, the mutant plants accumulated more than 10 times as much bacteria as wild type plants (wt = 3.93 log cfu/leaf disk std. 0.57, *at1g02130* = 5.22 std. 0.9). Hence, *At1g02130* contributes to disease resistance in *A. thaliana*. It is possible that the *At1g02130* mutation inhibits defense responses that are dependent upon SGT1 interactions.

15 The bait fragment encoding amino acid 200-368 of OsSGT1 was found to interact with Archain delta COP, PN28982. A BLAST analysis of the amino acid sequence of PN28982 indicated that this prey protein is 92% similar to rice archain delta COP (gi|2506139). Cytosolic coat proteins that bind reversibly to membranes have a central
20 function in membrane transport within the secretory pathway. One well-studied example is COPI or coatomer, a heptameric protein complex that is recruited to membranes by the GTP-binding protein Arf1. Assembly into an electron-dense coat then helps in budding off membrane to be transported between the endoplasmic reticulum (ER) and Golgi apparatus. Activated Arf1 brings coatomer to membranes. However, once associated
25 with membranes, Arf1 and coatomer have different residence times: coatomer remains on membranes after Arf1-GTP has been hydrolysed and dissociated. Rapid membrane binding and dissociation of coatomer and Arf1 occur stochastically, even without vesicle budding. This continuous activity of coatomer and Arf1 generates kinetically stable membrane domains that are connected to the formation of COPI-containing transport
30 intermediates. This role for Arf1/coatomer might provide a model for investigating the behaviour of other coat protein systems within cells. (Presley *et al.*, *Nature* 417(6885):

187-193, 2002). It is possible that this delta COP interacts with the OsSGT1 and a Ras GTPase to coordinate membrane transport for proteolytically processed proteins.

An *A. thaliana* homologue to PN28982 was identified by BLAST. At5g05010 shares 77% amino acid similarity with PN28982. To see if *Arabidopsis* homologues of PN28982 have roles in disease resistance *Arabidopsis thaliana* with T-DNA insertions in At5g05010 (line SAIL_84_C10) was identified from a random insertion seed library. DNA regions surrounding the insertions were sequenced and revealed that the T-DNAs were located within the promoter of At5g05010. Plants were backcrossed and plants homozygous for the T-DNA insertion were identified by PCR. Homozygous mutants and wild type plants were challenged with *Pseudomonas syringae* pv. *maculicola* ES4326 and plants were assayed for amount of *P. syringae* bacteria accumulation 3 days post inoculation (Glazebrook *et al.*, *supra*). These experiments were repeated twice on at least six plants. Data are reported as means and standard deviations of the log of colony forming units per leaf cm². By three days after inoculation, the mutant plants accumulated more than 10 times as much bacteria as wild type plants (wt = 3.93 log cfu/leaf disk std. 0.57, at5g05010 = 5.24 std. 0.52). Hence, At5g05010 contributes to disease resistance in *A. thaliana*. It is possible that the At5g05010 mutation inhibits defense responses that are dependent upon SGT1 interactions.

The bait fragment encoding amino acid 200-368 of OsSGT1 was found to interact with fibrillin-like protein, PN29042. A BLAST analysis of the amino acid sequence of OsPN29037 indicated that this prey protein is 75% similar to the potato fibrillin homolog CDSP34 precursor from chloroplasts (gi|7489242). Plastid lipid-associated proteins, also termed fibrillin/CDSP34 proteins, are known to accumulate in fibrillar-type chromoplasts such as those of ripening pepper fruit, and in leaf chloroplasts from Solanaceae plants under abiotic stress conditions. Further, substantially increased levels of fibrillin/CDSP34 proteins are shown in various dicotyledonous and monocotyledonous plants in response to water deficit. (Langenkamper *et al.*, *J. Exp. Bot.* 52(360): 1545-1554, 2001) In water-stressed tomato plants, similar increases in the CDSP 34-related transcript amount were noticed in wild-type and ABA-deficient flacca mutant, but protein accumulation was observed only in wild-type, suggesting a posttranscriptional role of

ABA in CDSP 34 synthesis regulation. Substantial increases in CDSP 34 transcript and protein abundances were also observed in potato plants subjected to high illumination. The CDSP 34 protein is proposed to play a structural role in stabilizing stromal lamellae thylakoids upon osmotic or oxidative stress. (Gillet *et al.*, *Plant J.* 16(2): 257-262, 1998).

5 A BLAST analysis comparing the nucleotide sequence of PN29042 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS011738 (100%) as the closest match. Gene expression experiments indicated that this gene is up-regulated by ABA treatment.

10 An *A. thaliana* homologue to PN29042 was identified by BLAST. At4g22240 shares 79% amino acid similarity with PN29042. To see if Arabidopsis homologues of PN29042 have roles in disease resistance *Arabidopsis thaliana* with T-DNA insertions in At4g22240 (line SAIL_691_B11) was identified from a random insertion seed library. DNA regions surrounding the insertions were sequenced and revealed that the T-DNAs were located within exon 1 of At4g22240. Plants were backcrossed and plants
15 homozygous for the T-DNA insertion were identified by PCR. Homozygous mutants and wild type plants were challenged with *Pseudomonas syringae* pv. *maculicola* ES4326 and plants were assayed for amount of *P. syringae* bacteria accumulation 3 days post inoculation (Glazebrook *et al.*, *supra*). These experiments were repeated twice on at least six plants. Data are reported as means and standard deviations of the log of colony
20 forming units per leaf cm². By three days after inoculation, the mutant plants accumulated more than 10 times as much bacteria as wild type plants (wt = 3.93 log cfu/leaf disk std. 0.57, at4g22240= 5.21 std. 0.43). Hence, At4g22240 contributes to disease resistance in *A. thaliana*. It is possible that the At4g22240 mutation inhibits defense responses that are dependent upon SGT1 interactions.

25 Additionally, the bait fragment encoding amino acid 200-368 of OsSGT1 was found to interact with HSP70-like protein, PN23949. A BLAST analysis of the amino acid sequence of OsPN3949 indicated that this prey protein is 71% similar to the cucumber 70K heat shock protein found in chloroplasts (gi|7441856). Heat shock
30 proteins (reviewed in Bierkens *et al.*, *Toxicology* 153(1-3): 61-72, 2000) are stress proteins that function as intracellular chaperones to facilitate protein folding/unfolding

and assembly/disassembly. They are selectively expressed in plant cells in response to a range of stimuli, including heat and a variety of chemicals. As regulators, HSP proteins are thus part of the plant protective stress response. A BLAST analysis comparing the nucleotide sequence of PN23949 against TMRI's GeneChip® Rice Genome Array
5 sequence database identified probeset OS015016 (97%) as the closest match. Gene expression experiments indicated that this gene is down-regulated by herbicide and JA treatment.

Yeast Two-Hybrid Using OsERP (PN20696) as Bait

10 Next, one of the proteins found to interact with OsSGT1, namely the elicitor responsive protein PN20696 (gi|11358958; OsERP), was used as a bait. As shown in Table 27, the rice elicitor responsive protein PN20696 (gi|11358958; OsERP) was found to interact with a receptor-like protein kinase like protein, PN31085. A BLAST analysis of the amino acid sequence of OsPN31085 indicated that this prey protein is 48% similar
15 to a rice receptor like protein kinase (gi|7434420). The receptor protein kinases include a large group of proteins and most contain a cytoplasmic protein kinase catalytic domain, a transmembrane region, and and/or an extracellular domain consisting of leucine-rich repeats, which are thought to interact with other macromolecules. Cell to cell communication is likely mediated by receptor kinases which have important roles in plant
20 morphogenesis.

OsERP was also found to interact with pyruvate orthophosphate dikinase, PN20674. A BLAST analysis of the amino acid sequence of PN20674 indicates that this prey protein is 97% similar to rice pyruvate orthophosphate dikinase (gi|743444). Pyruvate orthophosphate dikinase (PPDK) is known for its role in C4 photosynthesis but
25 has no established function in C3 plants. Absciscic acid, PEG and submergence were found to markedly induce a protein of about 97 kDa, identified by microsequencing as PPDK, in rice roots (C3). One rice PPDK is ABA-induced protein from roots. Western blot analysis showed a PPDK induction in roots of rice seedlings during gradual drying, cold, high salt and mannitol treatment, indicating a water deficit response. PPDK was
30 also induced in the roots and sheath of submerged rice seedlings, and in etiolated rice

seedlings exposed to an oxygen-free N₂ atmosphere, which indicated a low-oxygen stress response. None of the stress treatments induced PDK protein accumulation in the lamina of green rice seedlings. Pdk transcripts were found to accumulate in roots of submerged seedlings, concomitant with the induction of alcohol dehydrogenase 1. Low-oxygen stress triggered an increase in PDK activity in roots and etiolated rice seedlings, accompanied by increases in phosphoenolpyruvate carboxylase and malate dehydrogenase activities. The results indicate that cytosolic PDK is involved in a metabolic response to water deficit and low-oxygen stress in rice, an anoxia-tolerant species (Moons *et al.*, *Plant J.* **15**(1): 89-98, 1998).

10

Additionally, OsERP was found to interact with gamma adaptin, PN24292. A BLAST analysis of the amino acid sequence of PN24292 indicated that this prey protein is 97% similar to the Arabidopsis gamma adaptin (gi|5091510). Eukaryotic vesicular transport requires the recognition of membranes through specific protein complexes. The heterotetrameric adaptor protein complexes 1, 2, and 3 (AP1/2/3) are composed of two large, one small, and one medium adaptin subunit. Large subunits of AP1/2/3 are homologous and two subunits of the heptameric coatamer I (COPI) complex belong to this gene family. In addition, all small subunits and the aminoterminal domain of the medium subunits of the heterotetramers are homologous to each other; this also holds for two corresponding subunits of the COPI complex. AP1/2/3 and a substructure (heterotetrameric, F-COPI subcomplex) of the heptameric COPI have a common ancestral complex (called pre-F-COPI). Since all large and all small/medium subunits share sequence similarity, the ancestor of this complex is inferred to have been a heterodimer composed of one large and one small subunit. (Schledzewski *et al.*, *J. Mol. Evol.* **48**(6): 770-778, 1999). An archaean delta COP interacts with OsSGT1 which interacts with the Gamma adaptin bait ERP.

15

20

25

OsERP was also found to interact with xanthine dehydrogenase, PN29997. A BLAST analysis of the amino acid sequence of PN29997 indicated that this prey protein is 66% similar to the *Arabidopsis* xanthine dehydrogenase (gi|15236216). Xanthine dehydrogenase is the enzyme responsible for xanthine degradation. Xanthine

30

dehydrogenase is involved in purine catabolism and stress reactions. A BLAST analysis comparing the nucleotide sequence of PN29997 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS013724 (100%) as the closest match. Gene expression experiments indicated that this gene is expressed in seeds.

5

OsERP was also found to interact with ubiquitin specific protease, PN30843. A BLAST analysis of the amino acid sequence of PN30843 indicated that this prey protein is 40% similar to an Arabidopsis ubiquitin specific protease (gi|11993486). The ubiquitin/26S proteasome pathway is a major route for selectively degrading cytoplasmic and nuclear proteins in eukaryotes. In this pathway, chains of ubiquitins become attached to short-lived proteins, signaling recognition and breakdown of the modified protein by the 26S proteasome. During or following target degradation, the attached multi-ubiquitin chains are released and subsequently disassembled by ubiquitin-specific proteases (UBPs) to regenerate free ubiquitin monomers for re-use. T-DNA insertion mutations in an Arabidopsis ubiquitin protease cause an embryonic lethal phenotype, with the homozygous embryos arresting at the globular stage. The arrested seeds have substantially increased levels of multi-ubiquitin chains, indicative of a defect in ubiquitin recycling. Thus, there is essential role for the ubiquitin/26S proteasome pathway in general and for AtUBP14 in particular during early plant development (Doelling *et al.*, *Plant J.* 27(5): 393-405, 2001). SGT1 also interacts with components of the ubiquitin/26S proteasome pathway and the ERP that interacts with this ubiquitin specific protease interacts with OsSGT. This protease may have roles in disease resistance as well as development.

25 OsERP was also found to interact with pectinesterase, PN30845. A BLAST analysis of the amino acid sequence of PN30845 indicated that this prey protein is 71% similar to a rice pectinesterase (gi|15528783). Pectinesterases catalyze the esterification of cell wall polygalacturonans. In dicot plants, these ubiquitous cell wall enzymes are involved in important developmental processes including cellular adhesion and stem elongation. A BLAST analysis comparing the nucleotide sequence of PN30845 against
30 TMRI's GeneChip® Rice Genome Array sequence database identified probeset

OS007057 (99%) as the closest match. Gene expression experiments indicated that this gene is up-regulated as a result of JA treatment, high saline growth conditions and herbicide treatment.

5 OsERP was also found to interact with several proteins, namely PN30870, PN29984, PN30844, PN29983, PN30868 and PN30857. A BLAST analysis of the amino acid sequence of PN30870, PN29984, PN30844, PN29983, PN30868 and PN30857 indicates that these prey proteins have no sufficient homology to any other characterized proteins. However, based on association with the rice elicitor responsive
10 protein PN20696, these proteins may have roles in disease resistance or cell cycling.

A BLAST analysis comparing the nucleotide sequence of PN30857 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS008661.1 (99%) as the closest match. Gene expression experiments indicated that this
15 gene is up-regulated as a result of blast infection.

An *A. thaliana* homologue to PN29983 was identified by BLAST. At2g36950 shares 52% amino acid similarity with PN29983. To see if Arabidopsis homologues of PN29983 have roles in disease resistance, *Arabidopsis thaliana* with T-DNA insertions in
20 At2g36950 (line SAIL_779_E11) was identified from a random insertion seed library. DNA regions surrounding the insertions were sequenced and revealed that the T-DNAs were located within exon 3 of At2g36950. Plants were backcrossed and plants homozygous for the T-DNA insertion were identified by PCR. Homozygous mutants and wild type plants were challenged with *Pseudomonas syringae* pv. *maculicola* ES4326 and
25 plants were assayed for amount of *P. syringae* bacteria accumulation 3 days post inoculation (Glazebrook *et al.*, *supra*). These experiments were repeated twice on at least six plants. Data are reported as means and standard deviations of the log of colony forming units per leaf cm². By three days after inoculation, the mutant plants accumulated more than 10 times as much bacteria as wild type plants (wt = 3.94 log
30 cfu/leaf disk std. 0.57, at2g36950 = 5.95 std. 0.72). Hence, At2g36950 contributes to

disease resistance in *A. thaliana*. It is possible that the At2g36950 mutation inhibits defense responses that are dependent upon ERP/SGT1 interactions.

It should be noted that the all of the following bait proteins, namely OsSGT, ring
5 zinc finger, PN20115, and shaggy kinase, PN20621, identified proline rich protein, PN23221, as their prey. OsSGT and PN23221 have been described earlier in this Example.

A BLAST analysis of the amino acid sequence of ring zinc finger PN20115 indicated that this bait protein is 65% similar to *A. thaliana* ring zinc finger protein
10 At1g63170. The RING domain is a conserved zinc finger motif, which serves as a protein-protein interaction interface. This protein may interact with other proteins to control developmental or stress tolerance processes. A BLAST analysis comparing the nucleotide sequence of PN20115 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS015830 (90%) as the closest match. Gene
15 expression experiments indicated that this gene is up-regulated as a result of conditions of drought.

A BLAST analysis of the amino acid sequence of shaggy kinase PN20621 indicated that this bait protein is the rice shaggy kinase (gi|131677093). GSK3/SHAGGY is a highly conserved serine/threonine kinase implicated in many signaling pathways in
20 eukaryotes. Many GSK3/SHAGGY-like kinases have been identified in plants. The Arabidopsis BRASSINOSTEROID-INSENSITIVE 2 (BIN2) gene encodes a GSK3/SHAGGY-like kinase. Gain-of-function mutations within its coding sequence or its overexpression inhibit brassinosteroid (BR) signaling, resulting in plants that resemble BR-deficient and BR-response mutants. In contrast, reduced BIN2 expression via
25 cosuppression partially rescues a weak BR-signaling mutation. Thus, BIN2 acts as a negative regulator to control steroid signaling in plants (Li and Nam, *Science* 295(5558): 1299-1301, 2002).

Summary

30 As one of the major human staples, rice has been a target of genetic engineering for higher yields and resistance to diseases, pests, and environmental stresses of various

kinds. The proteins identified in the present Example have presumed roles in cell cycle processes and/or the stress response. Knowledge of the proteins and molecular interactions associated with cell cycle processes and stress response in rice could lead to important applications in agriculture. Modulation of these interactions may be exploited to effect changes in plant development or growth that would result in increased crop yield and tolerance to environmental stress conditions.

Plant disease response often mimics certain normal developmental processes. For example, plants responses to fungal gibberellic acid and fusaric acid toxin are similar to responses to plant-produced gibberellin and auxin, respectively (Hedden and Kamiya, *Annual Rev. Plant Physiol. Plant Mol. Biol.* 48: 431, 1977; Baunsgaard *et al.*, *Plant J.* 13: 661, 1998). The same can be said for abiotic stress responses and certain stages of plant development. Leaf cells undergoing dehydration stress express some of the same genes that embryonic cells express during development or seed desiccation (Medina *et al.*, *Plant Physiol.* 125: 1655, 2001). Since systematic regulation of gene expression drives developmental processes and stress responses (Chen *et al.*, *Plant Cell* 14: 559, 2002) it is likely that there is a broader overlapping set of genes and their cognate proteins involved in such responses. This Example describes one such overlapping set of genes.

The results described in this Example are useful for predicting gene function in rice or other plants. For example, rice has a homolog (OsSGT1; gb|AAF18438) to the barley SGT1 and *A. thaliana* SGT1b proteins that participate in pathogen defense through interactions with resistance gene and ubiquitinylation protein degradation pathways. *OsSGT1* is inducible by blast infection and likely participates in pathogen defense. *OsSGT1* interacted with several undefined and known proteins, including one whose transcript is induced upon treatment with a rice blast fungal elicitor (gb|AF090698). The elicitor-responsive protein (OsERP) interacted with other undefined proteins and an ubiquitin protease-related protein, which implicates OsERP in SGT1 mediated protein degradation. These rice proteins, as well as other plant homologs, are suspected to have associated roles in disease resistance.

A. thaliana proteins homologous to OsERP (PN20696), Ras GTPase (PN24063), Archain delta COP-like (28982), fibrillin-like (PN29042) and to one of the undefined proteins that interacted with OsERP (PN29983) have also been identified. *A. thaliana* homozygous for insertion mutations in the cognate genes were challenged with

5 *Pseudomonas syringae*. By three days after inoculation, the mutant plants accumulated more than 10 times as many bacteria as wild type plants. Hence, these Arabidopsis homologs contribute to disease resistance in *A. thaliana*. It is possible that these mutations inhibit defense responses that are dependent upon SGT1 interactions. Based upon homology and the interaction map, the rice homologs from which are associated the

10 Arabidopsis genes may also involved in disease resistance and other processes utilizing SGT1 as a factor. These results demonstrate that the combined datasets can be used to predict gene functions that can be verified using phenotypes of mutants.

Example VIII

This Example describes the identification and characterization of rice proteins that

15 interact at the cell wall in response to biotic stress. As has been described above, an automated, high-throughput yeast two-hybrid assay technology was used to identify proteins interacting with rice chitinase, class III, and with cellulose synthase catalytic subunit. The sequences encoding the protein fragments used in the search were then compared by BLAST analysis against proprietary and public databases to determine the

20 sequences of the full-length genes. The proteins found appear to be localized or targeted to the cell wall and to participate in the plant pathogen-induced defense response. The identification and characterization of proteins participating in pathways and biochemical reactions associated with defense against pathogens in rice may allow the development of genetically modified crops with enhanced or reduced disease resistance.

25

Chitinases are glycohydrolases that degrade chitin, a structural component of insects and plant pathogens such as nematodes, fungi, and bacteria. These enzymes are involved in multiple biological functions that include defense against chitin-containing pathogens, with class III chitinases having a substrate specificity for bacterial cell walls

30 (Brunner *et al.*, *Plant J.* 14(2): 225-34, 1998). Chitinase was chosen as a bait for these interaction studies based on its relevance to TMRI's plant health programs. The high

potential for specific enzyme-substrate interactions makes these proteins suitable for two-hybrid assays. The identification of rice genes encoding proteins involved in the plant response to pathogens are important to agriculture, as their discovery may allow genetic manipulation of crops to obtain plants with enhanced or reduced disease resistance.

5

The second bait used in this Example, namely cellulose synthase catalytic subunit, is part of a membrane-bound enzyme complex involved in the synthesis of cellulose, an essential component of the cell wall of higher plants whose production is central to morphogenesis and many other biological processes in plants (reviewed in Perrin R.M.,
10 *Curr. Biol.* 11(6): R213-R216, 2001).

This example provides newly characterized rice proteins interacting with a rice chitinase, class III (OsCHIB1), and with rice cellulose synthase catalytic subunit, RSW1-like (OsCS). An automated, high-throughput yeast two-hybrid assay technology
15 (provided by Myriad Genetics Inc., Salt Lake City, UT) was used to search for protein interactions with the chitinase and cellulose synthase bait proteins.

Results

Chitinase, class III, was found to interact with rice catalase A, an antioxidant
20 enzyme that is part of the plant's detoxification mechanism against molecules induced in response to environmental stresses. A second interactor, cellulose synthase catalytic subunit, is an enzyme involved in cellulose biosynthesis and is the second bait protein of this Example. The search also identified four novel rice proteins interacting with
25 chitinase: a protein similar to plant ABC transporter proteins, which play an important role in defense responses by eliminating toxins from tissues; a peptidase similar to *Arabidopsis thaliana* glutamyl aminopeptidase, whose proteolytic activity may be associated with activation of signaling molecules during the response of the plant to pathogens; a protein similar to a putative ATPase from *A. thaliana*, and one unknown
30 protein, similar to a putative protein from *A. thaliana*.

30

The cellulose synthase catalytic subunit bait clone was found to interact with itself and with twelve proteins. These include three known rice proteins: the DNAJ homologue, a type of molecule known to participate in the plant protective stress response as a regulator of heat shock proteins, and two proteins that function as membrane-spanning pumps: the product of the *salT* gene, which is induced by salt and stress, and the channel protein aquaporin. Nine interactors are novel proteins: a DNA-damage inducible-like protein with a putative role in the plant defense mechanism against nucleic acid damage; a putative BAG protein which presumably participates in the plant stress response by regulating heat shock proteins; a protein similar to the riboflavin precursor 6,7-dimethyl-8-ribityllumazine synthase precursor from *A. thaliana* and possibly involved in biosynthesis of riboflavin during oxidative stress; a protein similar to soybean calcium-dependent protein kinase and one similar to *A. thaliana* putative zinc finger protein, with likely roles as mediators of molecular signaling or transcription following damage to the cell wall; and four proteins of unknown function.

The interacting proteins of the Example are listed in Table 29 and Table 30 below, followed by detailed information on each protein and a discussion of the significance of the interactions. A diagram of the interactions is provided in Figure 5. The nucleotide and amino acid sequences of the proteins of the Example are provided in Figure 15.

Some of the proteins identified represent rice proteins previously uncharacterized. These proteins appear to participate in the plant defense mechanism against pathogens. Based on their presumed biological function and on their ability to specifically interact with the chitinase and cellulose synthase bait proteins, the interacting proteins may be localized or targeted to the cell wall, where they are involved in biochemical reactions and gene induction associated with local or systemic defense against pathogens.

Table 29. Interacting Proteins Identified for OsCHIB1 (Chitinase, Class III).

The Myriad names and the TMRI names of the clones of the proteins used as baits and found as preys are given. Nucleotide/protein sequence accession numbers for the proteins of the Example (or related proteins) are shown in parentheses under the protein name. The bait and prey coordinates (Coord) are the amino

acids encoded by the bait fragment(s) used in the search and by the interacting prey clone(s), respectively. The source is the library from which each prey clone was retrieved.

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (Source)
BAIT PROTEIN:			
OsCHIB1 PN19651	<i>O. sativa</i> Chitinase, Class III (AF296279; AAG02504)		
INTERACTORS:			
OsCATA PN20899	<i>O. sativa</i> Catalase A Isozyme (D29966; BAA06232)	10-200	332-433 (input trait)
OsCS* PN19707	<i>O. sativa</i> Cellulose Synthase Catalytic Subunit, RSW1-Like (AF030052; AAC39333)	10-200	411-489 (input trait)
OsPN22823	Novel Protein PN22823, Similar to ABC Transporter Proteins (T02187, AB043999.1, NP_171753; e=0)	10-200	25-106 (input trait)
OsPN22154	Novel Protein PN22154, Similar to <i>A. thaliana</i> Glutamyl Aminopeptidase (AL035525; e=0)	10-200	390-562 (input trait)
OsPN29041	Novel Protein PN29041, Fragment, Similar to <i>A. thaliana</i> Putative ATPase (AAG52137; e ⁻¹⁷)	10-200	2x 5-108 (input trait)
OsPN22020 (FL_R01_P005_C09.g.1a.Sp 6a)	Novel Protein PN22020, Fragment, Similar to <i>A. thaliana</i> Putative Protein (NP_197783; 3e ⁻³⁴)	10-200	3x 76-170 128-170 (input trait)

* The cellulose synthase catalytic subunit was also used as a bait; its interactions are shown in Table 30.

5 Table 30. Interacting Proteins Identified for OsCS (Cellulose Synthase Catalytic Subunit, RSW1-Like)

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (Source)
BAIT PROTEIN:			
OsCS PN19707	<i>O. sativa</i> Cellulose Synthase Catalytic Subunit, RSW1-Like (AF030052; AAC39333)		
INTERACTORS:			
OsCS	<i>O. sativa</i> Cellulose Synthase Catalytic Subunit, RSW1-Like (AF030052; AAC39333)	316-583	316-582 (input trait)
OsAAB53810 PN29086	<i>O. sativa</i> salT Gene Product (AF001395; AAB53810.1)	316-583	6-145 (output trait)
OsPIP2A PN29098	<i>O. sativa</i> Aquaporin (AF062393)	316-583	123-290 (output trait)

OsPN22825	Novel Protein PN22825, Fragment	316-583	5-129 (input trait)
OsPN29076	Novel Protein PN29076, Fragment	316-583	1-187 43-388 122-304 (output trait)
OsPN29077	Novel Protein PN29077, Fragment, Similar to <i>A. thaliana</i> DNA-Damage Inducible Protein DDI1-Like (BAB02792; 5e ⁻⁹⁴)	316-583	4x 1-242 (output trait)
OsPN29084	Novel Protein PN29084, Fragment, Similar to Soybean (<i>Glycine max</i>) Calcium-Dependent Protein Kinase (A43713, 2e ⁻⁷⁹)	316-583	3x 1-253 (output trait)
OsPN29113	<i>O. sativa</i> DNAJ Homologue (BAB70509.1)	316-583	1-92 (output trait)
OsPN29115	Novel Protein PN29115, Fragment, Similar to <i>A. thaliana</i> 6,7-Dimethyl-8-Ribityllumazine Synthase Precursor (AAK93590, 6e ⁻³⁷)	316-583	1-188 (output trait)
OsPN29116	Novel Protein PN29116, Fragment	316-583	1-169 (output trait)
OsPN29117 (FL_R01_P078_N11. fasta.contig1)*	Novel Protein PN29117	316-583	-7-151 (output trait)
OsPN29118	Novel Protein PN29118, Fragment	316-583	1-136 (output trait)
OsPN29119 (FL_R01_P084_PO 1.g.1a.Sp6a)	Novel Protein PN29119, Fragment	316-583	-53 to 155 (output trait)

* OsPN29117 also interacts with heat shock protein hsp70 (OsHSP70, PN20775): three prey clones of OsPN29117 (one encoding amino acids 11-160, two encoding amino acids 29-160) from the output trait library interacted with a clone (amino acids 138-360) of OsHSP70 used as bait.

5

Yeast Two-Hybrid Using OsCHIB1 (Chitinase, Class III) as Bait

The rice class III chitinase (GenBank Accession No. AF296279) is a 286-amino acid protein. Chitinases are glycohydrolases that degrade chitin. Chitin is a structural component of insects, nematodes, fungi, and bacteria. Chitinases are one of the several kinds of pathogenesis-related (PR) proteins induced in higher plants in response to infection by pathogens (reviewed in Stintzi *et al.*, *Biochimie*. 75(8): 687-706, 1993).

10

While chitinases perform multiple biological functions, the class III chitinases' substrate specificity for bacterial cell walls suggests a main role for these enzymes as defense proteins (Brunner *et al.*, *supra*). The enzyme directly attacks the pathogen by degrading the fungal or bacterial cell wall.

5 The bait fragment used in this search encodes amino acids 10 to 200 of OsCHIB1 (Chitinase, Class III). This region of the protein includes the active site of the enzyme (amino acids 127 to 135). There is no match for the gene encoding OsCHIB1 on TMRI's GeneChip® Rice Genome Array.

10 OsCHIB1 (Chitinase, Class III) was found to interact with OsCATA PN20899 (*O. sativa* Catalase A Isozyme (D29966; BAA06232)). Catalase A (GenBank Accession No. D29966) is the product of the rice CatA gene, which was identified by Higo and Higo, *Plant Mol. Biol.* 30(3): 505-521, 1996 as the homologue of the Cat-3 gene from Indian corn (*Zea mays*) (GenBank Accession No. L05934). Both rice CatA
15 and *Z. mays* Cat-3 genes belong to the monocot-specific group, one of three groups into which plant catalase genes have been classified based on their molecular evolution from a common ancestor (Guan and Scandalios, *J. Mol. Evol.* 42(5): 570-579, 1996). Rice catalase A contains 491 amino acids with two catalytic sites in position H65 and N138, and a heme binding-site in position Y348. The heme group is a cofactor for catalases'
20 enzymatic activity. Higo and Higo, *supra*, showed that the CatA gene is expressed at high levels in seeds during early development and also in young seedlings, and that this gene is induced by the herbicide paraquat, but not or only slightly by abscisic acid (ABA), wounding, salicylic acid, and hydrogen peroxide.

 Catalases are stress-induced enzymes found in almost all aerobic organisms.
25 They are part of the enzymatic detoxification mechanism against active oxygen species (AOS) in plant cells. AOS are induced in response to environmental stress and act as signaling molecules to activate multiple defense responses through induction of PR genes and of other signaling molecules (*e.g.*, salicylic acid, SA), leading to increased stress tolerance (Lamb and Dixon, *Ann. Rev. Plant Biol.* 48 (1): 251, 1997). AOS, however,
30 can also damage proteins, membrane lipids, DNA and other cellular components of the plant. The balance between these two diverging effects depends on the tight control of

cellular levels of AOS, which is achieved through a diverse battery of oxidant scavengers. Among these antioxidant molecules, catalases protect plant cells from the toxic effects of the AOS precursor hydrogen peroxide generated in the oxidative burst by converting it to dioxygen and water (reviewed in Dat *et al.*, *Redox Rep.* 6(1): 37-42, 2001).

OsCHIB1 (Chitinase, Class III) was found to interact with *O. Sativa* Cellulose Synthase Catalytic Subunit, RSW1-Like (OsCS) (PN19707). The prey clone found in our search, retrieved from the input trait library, encodes amino acids 411 to 489 of rice cellulose synthase catalytic subunit. This region of the 583-amino acid protein is C-terminal to the transmembrane domains and is predicted by amino acid sequence analysis to be on the cytoplasmic side of the plasma membrane.

Cellulose synthase is a membrane-bound enzyme complex comprising multiple isoforms. Cellulose synthase catalytic subunit (GenBank Accession No. AF030052) is involved in the synthesis of cellulose, a polysaccharide that is an essential component of the cell wall of higher plants. Cellulose imparts mechanical properties to plants which determine plant growth and cell shape, and its production impacts many aspects of plant biology. Most plants synthesize cellulose at the plasma membrane through the activity of cellulose synthase. As part of a structure called the rosette, the enzyme extends nascent cellulose chains by adding a sugar nucleotide precursor, and these chains then assemble into microfibrils that align in the same direction on the surface of the plasma membrane. This process seems to depend on a precise organization and orientation of the rosette (Perrin, R.M., *Curr. Biol.* 11(6): R213-6, 2001). A mutation in the *A. thaliana rsw1* gene that causes cellulose disassembly results in altered root morphogenesis (Baskin *et al.*, *Aust. J. Plant Physiol.* 19(4): 427-437, 1992), indicating that proper cellulose synthesis is critical to plant development and morphology. Arioli *et al.*, *Science* 279(5351): 717-720, 1998 showed that the *rsw1* gene in *A. thaliana* encodes a catalytic subunit of cellulose synthase. However, genetic and biochemical evidence now supports the concept that a family of genes encode the catalytic subunit of cellulose synthase in higher plants, with various members showing tissue-specific expression or being differentially expressed in response to various conditions. These topics are reviewed in Perrin, R.M., *supra*. These

authors indicate that the presence of many genes for the cellulose synthase catalytic subunit in plants suggests that multiple isoforms of cellulose synthase may be needed in the same cell for the formation of functional multimeric complexes, most likely dimers. In addition, many other polypeptides have been detected within the rosette whose identities have not been determined. Interaction studies aimed at identifying the proteins interacting with synthase may help elucidate the organization of the cellulose synthase rosette machinery and address some of the questions that still remain about the biosynthesis of cellulose. There is no match for the gene encoding OsCS on TMRI's GeneChip® Rice Genome Array.

Cellulose synthase catalytic subunit was also used as a bait protein. Its interactors are shown in Table 30 and discussed in later in this Example.

OsCHIB1 (Chitinase, Class III) was found to interact with Protein PN22823, which is similar to ABC Transporter Proteins (OsPN22823). Protein PN22823 is a 1239-amino acid protein that includes ten predicted transmembrane domains (amino acids 45 to 61, 154 to 170, 174 to 190, 253 to 269, 295 to 311, 671 to 687, 715 to 731, 794 to 810, 818 to 834, and 933 to 949) and two ATP/GTP-binding site motifs A (P-loops) (amino acids 383 to 390 and 1031 to 1038). A BLAST analysis against the Genpept database indicated that PN22823 shares 55% identity with Japanese goldthread (*Coptis japonica*) CjMDR1 (GenBank Accession No. AB043999.1; e=0.0). CjMDR1 is a multidrug resistance gene expressed in the rhizome, where alkaloids are highly accumulated compared to other organs (Yazaki *et al.*, *J. Exp. Bot.* 52(357): 877-9, 2001). Other proteins highly similar to PN22823 include *A. thaliana* putative ABC transporter (GenBank Accession No. T02187; e=0) and putative P-glycoprotein (GenBank Accession No. NP_171753; e=0). These types of proteins contain ATP-binding cassettes (ABC) and belong to a family that includes P-glycoprotein (P-gp) and multidrug resistance-associated protein 2 (MRP2) (reviewed by Fardel *et al.*, *Toxicology* 167(1): 37-46, 2001). ABC proteins are membrane-spanning proteins that transport a wide variety of compounds across biological membranes, including phospholipids, ions, peptides, steroids, polysaccharides, amino acids, organic anions, drugs and other xenobiotics.

In mammals, ABC transporters participate in the biliary elimination of exogenous compounds and xenobiotics, and their expression can be up-regulated by these toxins. The large number of ABC transporter protein family members identified in *A. thaliana* (129 according to Sanchez-Fernandez *et al.*, *J. Biol. Chem.* 276(32): 30231-30244, 2001), suggests an important role for these proteins in plants. In agreement with this notion, ABC transporters were among the immediate early genes found to be up-regulated in a tropical *japonica* rice cultivar (*Oryza sativa* cv. *Drew*) in response to jasmonic acid, benzothiadiazole, and/or blast infection (Xiong *et al.*, *Mol. Plant Microbe Interact.* 14(5): 685-692, 2001). This suggests that ABC proteins play a role in defense against toxins in plants as they do in mammals. Most of the ABC transporters characterized in plants to date have been localized in the vacuolar membrane and are considered to be involved in the intracellular sequestration of cytotoxins (reviewed in Leslie *et al.*, *Toxicology* 167(1): 3-23, 2001). Furthermore, plant ABC transporters appear to have a role equivalent to that of the mammalian ABC transporter in multidrug resistance, as shown in a study in which an ABC transporter protein was up-regulated in a *Nicotiana plumbaginifolia* cell culture following treatment with a close analog of the antifungal diterpene sclareol (Jasinski *et al.*, *Plant Cell* 13(5): 1095-107, 2001). MRP homologues isolated from *A. thaliana* (AtMRPs) are implicated in providing herbicide resistance to plants (Rea *et al.*, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49: 727-760, 1998). There is also evidence that ABC transporter proteins act as hormone transporters as they do in mammals. Specifically, a mutation in one of the ABC transporters in *A. thaliana*, AtMRP5, results in decreased root growth and increased lateral root formation possibly due to the inability of the mutant AtMRP5 to act as an auxin conjugate transporter Gaedeke *et al.*, *EMBO J.* 20(8): 1875-1887, 2001).

25 A BLAST analysis comparing the nucleotide sequence of Novel Protein PN22823 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS_ORF012127_at (e^{-145} expectation value) as the closest match. Gene expression experiments indicated that this gene is induced by the fungal pathogen *M. grisea*.

OsCHIB1 (Chitinase, Class III) was found to interact with protein PN22154, which is similar to *A. thaliana* Glutamyl Aminopeptidase (OsPN22154). OsPN22154 is a 173-amino acid protein fragment that is 65% identical to a protein from *A. thaliana* (GenBank Accession No. AL035525) described as a homologue of mouse
5 aminopeptidase (GenBank Accession No.U35646). The cDNA sequence of the *A. thaliana* aminopeptidase-like protein and the rice genome sequence (as a template) were used to generate a rice DNA sequence coding for a protein of 874 amino acids, which is 54.7 % identical to the *A. thaliana* aminopeptidase-like protein. Indeed, domain analysis of the novel rice protein detected a peptidase M1 domain (amino acids 17 to 402), and a
10 zinc-binding domain (amino acids 311 to 320), suggesting that this protein is a metallo-aminopeptidase. It is unclear whether this protein is encoded by an orthologue or an analogue of the *A. thaliana* aminopeptidase-like gene. A BLAST analysis comparing the nucleotide sequence of Novel Protein PN22154 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS_004263_at ($4e^{-83}$ expectation
15 value) as the closest match. Gene expression experiments indicated that this gene is expressed in panicle.

OsCHIB1 (Chitinase, Class III) was found to interact with protein PN29041 (OsPN29041). A BLAST analysis indicated that this protein fragment is similar to putative ATPase from *A. thaliana* (GenBank Accession No. AAG52137; e^{-17}). ATPases
20 can be localized to the plasma membrane which is adjacent to the cell wall. There is no match for this gene on TMRI's GeneChip® Rice Genome Array, and thus no gene expression data that would allow prediction of its function during stress or infection. It is possible that this protein may have no role in pathogen invasion. However, it is part of the chitinase multiprotein complex identified in this Example through the yeast two-
25 hybrid interactions, which we suggest exists at the cell wall interface. One hypothesis is that the ATPase-like protein may reside in the plasma membrane and participate in cell wall synthesis. Further interaction data may help elucidate the biological significance of its participation in the chitinase multiprotein complex.

OsCHIB1 (Chitinase, Class III) was found to interact with protein PN22020
30 (OsPN22020). Protein PN22020 is a 175-amino acid protein fragment that shares 55%

identity with *A. thaliana* putative protein (GenBank Accession No. NP_197783; $3e^{-34}$). Analysis of the amino acid sequence identified a C2 domain (amino acids 5 to 90, $e=0.037$), as found in protein kinase C isozymes, which suggests that PN22020 may participate in signaling pathways similar to those modulated by protein kinase C.

5 Perhaps its interaction with chitin represents a signaling event that occurs in response to pathogen or toxin exposure. However, this domain has been detected in other kinases and nonkinase proteins (Ponting and Parker, *Protein Sci.* 5(1): 162-166, 1996). Identification of the full amino acid sequence of novel protein PN22020 may make it possible to determine the class of C2 domain-containing proteins to which it belongs.

10 A BLAST analysis comparing the nucleotide sequence of Novel Protein PN22020 against TMRI's GeneChip[®] Rice Genome Array sequence database identified probeset OS008182_r_at (e^{-102} expectation value) as the closest match. Gene expression experiments indicated that this gene is constitutively expressed in leaves, stems, roots, seeds, panicle and pollen.

15

Yeast Two-Hybrid Using OsCS as Bait

A second bait, namely *O. sativa* Cellulose Synthase Catalytic Subunit, RSW1-Like (OsCS; PN19707; GenBank Accession No. AF030052), was also used. This protein is described earlier in this Example because it was found to interact with the bait protein

20 *O. sativa* Chitinase, Class III (OsCHIB1; PN19651). The bait fragment used in the search encodes amino acids 316 to 583 of OsCS.

OsCS was found to interact with *O. sativa* Cellulose Synthase Catalytic Subunit, RSW1-like (OsCS). In other words, OsCS was found to interact with itself. The prey

25 clone was retrieved from the input trait library, and encoded almost the same amino acids as the bait clone (the prey clone encoded amino acids 316 to 582). The self-interaction supports the concept of cellulose synthase acting as a dimer, as has been suggested (see Perrin, R.M., *Curr. Biol.* 11(6): R213-R216, 2001)).

30 OsCS was also found to interact with *O. sativa* salT Gene Product (OsAAB53810). A BLAST analysis of the 145-amino acid protein OsAAB53810 amino

acid sequence indicated that this protein is the rice salT Gene Product (AAB53810.1; 100% identity; $3e^{-80}$). This protein is encoded by a cDNA clone, salT, which was isolated from rice roots subjected to salinity stress, as reported by Claes *et al.* (*Plant Cell* 2(1): 19-27, 1990). These authors showed that the salT mRNA is specifically expressed in sheaths
5 and roots from mature plants and seedlings in response to salt stress and drought. Expression data reported previously by Garcia *et al.*, *Planta* 207(2): 172-80, 1998 indicate that expression of salT in each region of the plant is dependent on the metabolic activity of the cells as well as on whether or not they are responding to stress. These authors also found that the salT gene is induced by gibberellic acid and abscisic acid and
10 suggest that induction by these growth regulators occurs through independent and possibly antagonistic pathways. Analysis of the OsAAB53810 protein sequence predicted a jacalin-like lectin domain (amino acids 14 to 145, $2.3e^{-32}$). Jacalin interacts with carbohydrates in a highly specific manner (Sankaranarayanan *et al.*, *Nat. Struct. Biol.* 3(7): 596-603, 1996).

15 OsCS was also found to interact with Aquaporin (OsPIP2a). Aquaporin (GenBank Accession No. AF062393) is a 290-amino acid protein that includes six predicted transmembrane domains (amino acids 48 to 64; 83 to 99, 131 to 147, 175 to 191, 207 to 223, and 254 to 270) and a Major Intrinsic Protein (MIP) family signature
20 (amino acids 34 to 271), as determined by amino acid sequence analysis. The prey clone retrieved from the output trait library encodes amino acids 123 to 290 of OsPIP2a, a region that includes the four most C-terminal predicted transmembrane domains and part of the MIP family signature. Aquaporin is thought to be a plasma membrane intrinsic protein (Malz and Sauter, *Plant Mol. Biol.* 40(6): 985-995, 1999). Such proteins facilitate
25 movement of small molecules, often times functioning as water channels. This is why OsPIP2a is also called aquaporin. Malz and Sauter identified OsPIP2a along with OsPIP1a and report that these two proteins possess several hallmark motifs and homologies that justify their assignment to their respective PIP subfamilies. They report that OsPIP2a and OsPIP1a display similar, but not identical, expression patterns in rice,
30 both being expressed at higher levels in seedlings than in adult plants, and that expression in the primary root is regulated by light. Furthermore, their study indicates that

gibberellic acid also regulates the expression of these OsPIP transcripts in internodes of deepwater rice plants induced to grow rapidly by submergence, although expression did not correlate with growth. In *A. thaliana*, different PIP proteins are expressed in response to different agonists and conditions, e.g., salt stress induces tonoplast intrinsic protein (SITIP), as reported by Pih et al., *Mol. Cells* 9(1): 84-90, 1999. These authors suggest that PIP proteins may be responsible for osmoregulation in plants under high osmotic stress such as a high salt condition.

OsCS was also found to interact with protein PN22825 (OsPN22825).
OsPN22825 is a 229-amino acid protein fragment for which the complete sequence is not known. A BLAST analysis against the public and Myriad's proprietary databases indicated that OsPN22825 is similar to two unknown proteins from *A. thaliana* (GenBank Accession No. NP_188565, 67% identity, $3e^{-82}$; and GenBank Accession No. AB025624, 37% identity, $3e^{-82}$). There is no match for the gene encoding OsPN22825 on TMRI's GeneChip® Rice Genome Array, and thus no gene expression data that would allow prediction of its function during stress or infection.

OsCS was also found to interact with protein PN29076 (OsPN29076).
OsPN29076 is a 389-amino acid protein fragment for which the complete sequence is not known. Analysis of the available amino acid sequence identified a cytochrome c family heme-binding site (amino acids 142 to 147). A BLAST analysis revealed no proteins with high similarity to OsPN29076, the best hit being an *A. thaliana* unknown protein (GenBank Accession No. AAF24616, 34% identity, $3e^{-46}$). Three prey clones encoding amino acids 1 to 187, 42 to 389, and 121 to 304 of OsPN29076 were retrieved from the output trait library. The clones share an overlapping region which spans amino acids 121 to 187 of OsPN29076 and which includes the cytochrome c family heme-binding site. There is no match for the gene encoding OsPN29076 on TMRI's GeneChip® Rice Genome Array, and thus no gene expression data that would allow prediction of its function during stress or infection. The lack of information about OsPN29076 makes it difficult to determine its function. Identification of the complete amino acid sequence for

OsPN29076 may contribute to clarifying the function of this protein and the biological significance of the OsCS-OsPN29076 interaction.

OsCS was also found to interact with protein PN29077, which is similar to *A. thaliana* DNA-Damage Inducible Protein DDI1-Like (OsPN29077). OsPN29077 is 243-amino acid protein fragment for which the complete sequence is not known. A BLAST analysis indicated that OsPN29077 shares 73% identity with *A. thaliana* DNA-damage inducible protein DDI1-like (GenBank Accession No. BAB02792; 5e⁻⁹⁴). DDI1 is thought to be a cell-cycle checkpoint protein in yeast and its expression is induced by a variety of DNA-damaging agents. Such proteins arrest cells at certain stages and regulate the transcriptional response to DNA damage (Zhu and Xiao, *Nucleic Acids Res.* 26(23): 5402-5408, 1998). DDI1 has been reported to interact with ubiquitin (Bertolaet *et al.*, *Nat. Struct. Biol.* 8(5): 417-422, 2001), an observation that supports the use of the yeast two-hybrid approach to study such proteins.

A BLAST analysis comparing the nucleotide sequence of OsPN29077 against TMRI's GeneChip[®] Rice Genome Array sequence database identified probeset OS016688.1_at (e⁻⁸³ expectation value) as the closest match. Gene expression experiments indicated that this gene is not specifically expressed in several different tissue types and is not specifically induced by a broad range of plant stresses, herbicides, and applied hormones.

OsCS was also found to interact with protein PN29084, which is similar to *G. max* calcium-dependent protein kinase (OsPN29084). OsPN29084 is a 284-amino acid protein fragment for which the complete sequence is not known. Analysis of the available amino acid sequence identified four EF-hand calcium-binding domains (amino acids 110 to 122, 146 to 158, 182 to 194, and 216 to 228). In agreement with the presence of these domains, a BLAST analysis indicated that OsPN29084 is highly similar to many calcium-dependent protein kinases including soybean (*G. max*) calcium-dependent protein kinase (GenBank Accession No. A43713, 81% identity, 2e⁻⁷⁹). This soybean protein also includes four EF-hand calcium-binding domains and requires calcium but not calmodulin or phospholipids for activity (Harper *et al.*, *Science*

252(5008): 951-954, 1991). Calcium can function as a second messenger through stimulation of such calcium-dependent protein kinases.

A BLAST analysis comparing the nucleotide sequence of OsPN29084 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset

5 OS004083.1_at (e^{-83} expectation value) as the closest match. Gene expression experiments indicated that this gene is not specifically expressed in several different tissue types and is not specifically induced by a broad range of plant stresses, herbicides, and applied hormones.

10 OsCS was also found to interact with *O. sativa* DNAJ homologue (OsPN29113). OsPN29113 is a 92-amino acid protein whose sequence includes an ATP/GTP-binding site motif A (P-loop, amino acids 43 to 50). A BLAST analysis of the available amino acid sequence indicated that OsPN29113 is the rice DNAJ homologue (Accession # BAB70509.1; 100% identity; $5e^{-39}$). In eukaryotic cells, DnaJ-like proteins regulate the
15 chaperone (protein folding) function of Hsp70 heat-shock proteins through direct interaction of different Hsp70 and DnaJ-like protein pairs (Cyr *et al.*, *Trends Biochem. Sci.* 19(4): 176-181, 1994). Heat shock proteins (reviewed in Bierkens, J.G., *Toxicology* 153(1-3): 61-72, 2000) are stress proteins that function as intracellular chaperones to facilitate protein folding/unfolding and assembly/disassembly. They are selectively
20 expressed in plant cells in response to a range of stimuli, including heat and a variety of chemicals. As regulators of heat shock proteins, DnaJ-like proteins are thus part of the plant protective stress response.

A BLAST analysis comparing the nucleotide sequence of OsPN29113 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset

25 OS002926_at (e^{-124} expectation value) as the closest match. Gene expression experiments indicated that this gene is not specifically expressed in several different tissue types and is not specifically induced by a broad range of plant stresses, herbicides, and applied hormones.

30 OsCS was also found to interact with protein PN29115, which is similar to *A. thaliana* 6,7-dimethyl-8-ribityllumazine synthase precursor (OsPN29115). OsPN29115

is a 188-amino acid protein fragment for which the complete sequence is not known. The available sequence includes an ATP/GTP-binding site motif A (P-loop, amino acids 94 to 101) and a 6,7-dimethyl-8-ribityllumazine synthase family signature (amino acids 42 to 186), as determined by analysis of the available amino acid sequence. The presence of
5 the latter domain is in agreement with the results of a BLAST analysis indicating that OsPN29115 shares 50% identity with *A. thaliana* putative 6,7-dimethyl-8-ribityllumazine synthase precursor (GenBank Accession No. AAK93590, $6e^{-37}$). The cofactor riboflavin is synthesized from the precursor 6,7-dimethyl-8-ribityllumazine (Nielsen *et al.*, *J. Biol. Chem.* **261**(8): 3661-3669, 1986). Flavins are involved in numerous biological processes
10 (reviewed by Massey, V., *Biochem. Soc. Trans.* **28**(4): 283-296, 2000). For example, they participate in electron transfer reactions and thereby contribute to oxidative stress through their ability to produce superoxide, but at the same time flavins participate in the reduction of hydroperoxides, the products of oxygen-derived radical reactions. Flavins also contribute to soil detoxification and are linked to light-induced DNA repair in plants.
15 The chemical versatility of flavoproteins is controlled by specific interactions with the proteins with which they are bound.

A BLAST analysis comparing the nucleotide sequence of OsPN29115 against TMRI's GeneChip[®] Rice Genome Array sequence database identified probeset OS015577_at (e^{-41} expectation value) as the closest match. Gene expression experiments
20 indicated that this gene is not specifically expressed in several different tissue types and is not specifically induced by a broad range of plant stresses, herbicides, and applied hormones.

OsCS was also found to interact with protein PN29116 (OsPN29116).
25 OsPN29116 is a 170-amino acid protein fragment for which the complete sequence is not known. Analysis of the available amino acid sequence identified a WD40 domain (amino acids 82 to 118), which is reported to participate in protein-protein interactions (Ajuh *et al.*, *J. Biol. Chem.* **276**(45): 42370-42381, 2001). A BLAST analysis indicated that OsPN29116 shares identity with two unknown proteins from *A. thaliana* (GenBank
30 Accession No. T45879, 67% identity, e^{-64} ; and GenBank Accession No. NP_181253, 69% identity, e^{-58}). The lack of information about OsPN29116 makes it difficult to

determine its function. Identification of the complete amino acid sequence for OsPN29116 may clarify the function of this protein and the biological relevance of the OsCSC-OsPN29116 interaction.

5 A BLAST analysis comparing the nucleotide sequence of OsPN29116 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS016500_r_at (e^{-12} expectation value) as the closest match. The expectation value is too low for this probeset to be a reliable indicator of the gene expression of OsPN29116.

OsCS was also found to interact with protein PN29117 (OsPN29117).
 10 OsPN29117 is a 237-amino acid protein that includes a ubiquitin domain (amino acids 12 to 84). Analysis of the amino acid sequence identified a BAG domain (amino acids 106 to 187, $2.1e^{-11}$), which is known to bind and regulate Hsp70/Hsc70 molecular chaperones (Briknarova *et al.*, *Nat. Struct. Biol.* 8(4): 349-352, 2001). The BAG family of
 15 cochaperones functionally regulates signal-transducing proteins and transcription factors important for cell stress responses, apoptosis, proliferation, cell migration and hormone action (Briknarova *et al.*, *supra*; Antoku *et al.*, *Biochem. Biophys. Res. Commun.* 286(5): 1003-1010, 2001). A BLAST analysis indicated that OsPN29117 shares identity with an *A. thaliana* unknown protein (GenBank Accession No. AAC14405, 44% identity, $4e^{-52}$).
 20 In agreement with the notion that OsPN29117 is a member of the BAG family of proteins, it was also found to interact with hsp70 (OsHSP70) (see note * under Table 30). Heat shock proteins (discussed above) are stress proteins which function as ATP-dependent intracellular chaperones and which are selectively expressed in plant cells in response to a range of stimuli, including heat and a variety of chemicals. As a regulator of heat shock proteins, the BAG protein OsPN29117 may thus be part of the plant
 25 protective stress response.

The prey clone retrieved in the search encodes amino acids 1 to 151 of OsPN29117, a region that includes the ubiquitin domain. Note that the prey clone includes a small portion (-7 to 0) of the 5' untranslated region, and thus its coordinates are shown in Table 2 as amino acids -7 to 151. A BLAST analysis comparing the
 30 nucleotide sequence of OsPN29117 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS017803_at (e^{-73} expectation value) as the closest

match. Gene expression experiments indicated that this gene is not specifically expressed in several different tissue types and is not specifically induced by a broad range of plant stresses, herbicides, and applied hormones.

OsCS was also found to interact with protein PN29118 (OsPN29118). OsPN29118 is a 136-amino acid protein fragment for which the complete sequence is not known. A BLAST analysis indicated that OsPN29118 has only weak similarity to proteins in the public domain and in Myriad's proprietary database, the best hit being an *A. thaliana* putative zinc finger protein SHI-like (GenBank Accession No. NP_201436, 42% identity, $5e^{-15}$). The protein with the next highest identity is an *A. thaliana* hypothetical protein (GenBank Accession No. T04595, 38% identity, $9e^{-15}$). Discovery of the complete amino acid sequence for OsPN29118 may contribute to clarifying the function of this protein and the biological relevance of the OsCSC-OsPN29118 interaction.

15 A BLAST analysis comparing the nucleotide sequence of OsPN29118 against
TMRI's GeneChip® Rice Genome Array sequence database identified probeset
OS004996.1_at (e^{-38} expectation value) as the closest match. Gene expression
experiments indicated that this gene is not specifically expressed in several different
tissue types and is not specifically induced by a broad range of plant stresses, herbicides,
20 and applied hormones.

OsCS was also found to interact with protein PN29119 (OsPN29119). OsPN29119 is a 327-amino acid protein fragment for which the complete sequence is not known. A BLAST analysis indicated that OsPN29119 shares 38% identity with an *A. thaliana* unknown protein, T17H3.9 (GenBank Accession No. AAD45997, 7e⁻⁵⁴). Discovery of the complete amino acid sequence for OsPN29119 may contribute to clarifying the function of this protein and the biological relevance of the OsCSC-OsPN29119 interaction. One prey clone encoding amino acids 1 to 155 of OsPN29119 was retrieved from the output trait library. This prey clone includes a portion of the 5' untranslated region and thus its coordinates are shown in Table 2 as amino acids -53 to 155. A BLAST analysis comparing the nucleotide sequence of OsPN29119 against

TMRI's GeneChip[®] Rice Genome Array sequence database identified probeset OS014829.1_at (e^{-131} expectation value) as the closest match. Gene expression experiments indicated that this gene is not specifically expressed in several different tissue types and is not specifically induced by a broad range of plant stresses, herbicides, and applied hormones.

Summary

Proteins that Interact with OsCHIB1 (Chitinase, Class III).

The yeast two-hybrid assay designed to search for proteins interacting with the chitinase bait proteins led to the isolation of proteins that appear to be associated with the plant defense response to pathogens. Resistance to disease occurs on several levels that include local and nonspecific systemic responses. The hypersensitive response (HR) in plants is a mechanism of local resistance to pathogenic microbes characterized by a rapid and localized tissue collapse and cell death at the infection site, resulting in immobilization of the intruding pathogen. This process is triggered by pathogen elicitors and orchestrated by an oxidative burst, which occurs rapidly after the attack (Lamb and Dixon, *Ann. Rev. Plant Biol.* 48(1): 251, 1997). The accumulation of active oxygen species (AOS) is a central theme during plant responses to both biotic and abiotic stresses. AOS are generated at the onset of the HR and might be instrumental in killing host tissue during the initial stages of infection. AOS also act as signaling molecules that induce expression of PR genes and production of other signaling molecules which participate in the signal cascade that leads to PR gene induction. The triggering of defense genes may extend to the uninfected tissues and the whole plant, leading to local resistance (LR) and systemic acquired resistance (SAR) (reviewed in Martinez *et al.*, *Plant Physiol.* 122(3): 757-766, 2000). As a result of SAR, other portions of the plant are provided with long-lasting protection against the same and unrelated pathogens.

Hydrogen peroxide from the oxidative burst plays an important role in the localized HR not only by driving the cross-linking of cell wall structural proteins, but also by triggering cell death in challenged cells and as a diffusible signal for the induction in adjacent cells of genes encoding cellular protectants such as glutathione S-transferase and glutathione peroxidase, and for the production of salicylic acid (SA). SA is thought

to act as a signaling molecule in LR and SAR through generation of SA radicals, a likely by-product of the interaction of SA with catalases and peroxidases, as reported by Martinez *et al.* (*supra*). These authors showed that recognition of a bacterial pathogen by cotton triggers the oxidative burst that precedes the production of SA in cells undergoing the HR, and that hydrogen peroxide is required for local and systemic accumulation of SA, thus acting as the initiating signal for LR and SAR. The involvement of catalase in SA-mediated induction of SAR in plants was previously demonstrated by Chen *et al.*, *Science* 262(5141): 1883-1886, 1993 who showed that binding of catalase to SA results in inhibition of catalase activity, and that consequent accumulation of hydrogen peroxide induces expression of defense-related genes associated with SAR. In this study, chitinase was found to interact with catalase A. Given the established role of chitinase as a defense protein, this interaction is consistent with the presence of the stress-induced catalase during pathogen attack and suggests that both enzymes may be located at the cell wall, where they participate in PR gene induction. The significance of the chitinase-catalase interaction as part of the defense response against microbes finds further support in the observation that fungal catalase has a role in protecting necrotrophic fungi from the deleterious effects of AOS during colonization of a host expressing the HR (Mayer *et al.*, *Phytochemistry* 58(1): 33-41, 2001). These organisms were shown to secrete catalase, among other enzymes, to remove or inactivate AOS from the host.

In addition, the cell wall may play a role in defense against bacterial and fungal pathogens by receiving information from the surface of the pathogen from molecules called elicitors, and by transmitting this information to the plasma membrane of plant cells, resulting in gene-activated processes that lead to resistance. One type of biochemical reaction induced by elicitors and associated with the hypersensitive response is the synthesis and accumulation of phytoalexins, antimicrobial compounds produced in the plant after fungal or bacterial infection (reviewed in Hammerschmidt, R., *Ann. Rev. Phytopathol.* 37: 285-306, 1999). One of the proteins found to interact with chitinase is an ABC transporter. ABC transporters are known to sequester cytotoxins, metabolites and other molecules from plant tissues. It is thus likely that the ABC transporter found to interact with chitinase resides at the cell wall, where it participates in the transport of toxins. Though the function of phytoalexins in the plant defense response has not been

thoroughly elucidated (Hammerschmidt, R., *supra*), it is tempting to speculate that the ABC transporter may be involved in the elimination of these toxins from the plant cells during the plant pathogen-induced defense response. Furthermore, gene expression experiments indicated that the gene encoding the ABC transporter protein is induced by the fungal pathogen *M. grisea*. These results are consistent with the putative role of this protein in the defense response induced by pathogenic fungi and bacteria in rice.

Chitinase was also found to interact with novel protein PN22154 similar to *A. thaliana* glutamyl aminopeptidase. While the specific function of this prey protein has not been determined, it is well known that proteolytic activity is a common component of plant defense mechanisms against pathogens. These mechanisms include both chitinases and proteases. Peptidase activity has been associated with regulation of signaling. Carboxypeptidases, for instance, hydrolytically remove the pyroglutamyl group from peptide hormones, thereby activating these signaling molecules. A carboxypeptidase regulates Brassinosteroid-insensitive 1 (BRI1) signaling in *A. thaliana* by proteolytic processing of a protein (Li *et al.*, *Proc. Natl. Acad. Sci. USA* **98**(10): 5916-5921, 2001). Based on its ability to interact with chitinase and on the well-established role of the latter in PR defense, chitinase and novel protein PN22154 may interact as components of a complex with chitinolytic and proteolytic activities targeted against plant invaders, and that the rice glutamyl aminopeptidase-like protein may have a role in activating signaling molecules at the cell wall that are involved in the plant defense response.

A fourth interactor found for chitinase is cellulose synthase catalytic subunit. This enzyme acts as a complex at the plasma membrane where it participates in cell wall synthesis, and its regulation may allow the plant to respond with morphological changes to physical insult produced by pathogen attack. This interaction may be significant to maintaining the balance of the metabolism of cell wall components during the defense response. It is possible that either chitinase resides at the cell wall where it interacts with cellulose synthase immediately following pathogen attack, or chitinase is targeted to this site and interacts with synthase after PR gene induction.

Aside from novel proteins PN22020 and PN29041, the rice proteins found to interact with chitinase appear to be localized at or recruited to the cell wall where they participate in the plant defense response to pathogen attack. Two of the interactors, an ABC transporter and a glutamyl aminopeptidase-like protein, are newly characterized proteins in rice.

As a whole, all of these proteins may interact as a multicomponent complex at the cell wall interface in the plant cell, and all may have roles in controlling AOS levels, inducing PR genes, and synthesizing and maintaining the integrity of the cell wall to protect the plant against the effects of pathogen invasion.

Proteins that Interact with Cellulose Synthase Catalytic Subunit (OsCS)

The interactions involving OsCS expand the stress-response protein network identified for the chitinase bait protein. OsCS interacts with several proteins that appear to participate in the plant response to pathogen-induced stress at the cell wall. Published evidence links some of these proteins to the plant response to various stresses. These include aquaporin (OsPIP2a) and salt-stress induced protein (OsAAB53810), two molecules that, although they may not have a direct role in disease resistance, can function as membrane-spanning pumps in the protein complex at the cell wall to regulate turgor pressure or transmit solutes. Moreover, the presence of the jacalin-like lectin domain in OsAAB53810 is of particular interest in the context of its interaction with an enzyme that synthesizes carbohydrate chains. Given the carbohydrate-binding property of jacalin (Sankaranarayanan *et al.*, *Nat. Struct. Biol.* 3(7): 596-603, 1996), OsAAB53810 may specifically bind nascent cellulose chains as they are produced by OsCS, thus playing an active role in OsCS-dependent events relating to cell wall metabolism. The fact that OsAAB53810 is induced by salt and stress supports a role for this protein in such physiological events.

Another interactor, the rice DNAJ homologue OsPN29113, likely participates in the plant protective stress response by regulating the chaperone function of heat shock proteins, which are induced by various forms of stress. It is possible that the interaction

of the DNAJ protein with cellulose synthase is part of the plant response to chemicals produced by pathogens or generated in cells undergoing the HR, and that such response is associated with injury to the cell wall that has occurred in response to the stress.

- 5 Among the novel proteins found to interact with OsCS, OsPN29077 is similar to *A. thaliana* DNA-damage inducible protein DDI1-like. Based on the expression of yeast DDI1 in response to DNA damage and on sequence homology, we speculate that OsPN29077 performs the same function as DDI1 and that the OsCS-OsPN29077 interaction is associated with the plant defense mechanism against DNA damage.
- 10 Likewise, we attribute the BAG-like protein OsPN29117 a putative role in the plant protective stress response as a regulator of heat shock proteins. In agreement with this role, OsPN29117 also interacts with hsp70, which our gene expression experiments indicate is expressed constitutively and is down-regulated by jasmonic acid (see chart in Appendix 1), a component of plant defense response pathways. Since OsPN29077 and
- 15 OsPN29117 interact with the cellulose synthase catalytic subunit, and the latter interacts with the pathogen-induced defense protein chitinase, these interactors may be a part of the same complex at the cell wall where they participate in the response to pathogen attack.
- 20 The novel protein OsPN29115 is similar to the riboflavin precursor 6,7-dimethyl-8-ribityllumazine synthase precursor from *A. thaliana*. Among the roles reported for riboflavin is its association with the redox reactions occurring as a result of oxidative stress (Massey, V., *Biochem. Soc. Trans.* **28**(4): 283-96, 2000). Based on this evidence and on sequence homology for the identified interactor, the OsCS-OsPN29115
- 25 interaction may link the plant response to stress and toxins produced by pathogens with structural changes requiring OsCS activity.

- Additional novel proteins interacting with OsCS include a protein similar to soybean calcium-dependent protein kinase (OsPN29084) and a protein similar to *A.*
- 30 *thaliana* putative zinc finger protein (OsPN29118). The similarities of these interactors to protein kinases and zinc finger proteins suggest that they function as mediators of

molecular signaling and transcription, respectively. Their interactions with OsCS may represent signaling or transcriptional events occurring after disruption following damage to the cell wall by pathogens, and these prey proteins may move from the cell wall to other parts of the cell to mediate such events. The OsCS-OsPN29084 interaction likely
5 represents a step in the transduction of an extracellular signal that results in a physiological response, while the OsCS-OsPN29118 interaction may be associated with transcriptional regulation also in response to an extracellular signal. This signal may be in the form of an insult to the plant produced by pathogen attack.

10 For the remaining proteins found to interact with OsCS—OsPN22825, OsPN29076, OsPN29116, and OsPN29119—based on their association with cellulose synthase and chitinase, these prey proteins may also be important factors for pathogen defense, cell wall integrity, or for holding together protein complexes.

15 Thus, the results presented in this Example show that proteins interacting with the cellulose synthase catalytic subunit are also part of the chitinase multiprotein complex localized at the cell wall interface.

Example IX

20 Janssens and Goris teach that type 2A serine/threonine protein phosphatases (PP2A) are important regulators of signal transduction, which they affect by dephosphorylation of other proteins (Janssens and Goris, *Biochem J.* **353**(Pt 3): 417-439, 2001). Members of the protein phosphatase 2A (PP2A) family of serine/threonine phosphatases contain a well-conserved catalytic subunit, the activity of which is highly
25 regulated (Janssens and Goris, *supra*). There are multiple PP2A isoforms in plants and other organisms, and they appear to be differentially expressed in various tissues and at different stages of development (Arino *et al.*, *Plant Mol. Biol.* **21**(3): 475-485, 1993). Harris *et al.* cites a number of reports describing the association of PP2A subunits with a variety of cellular proteins in addition to regulatory subunits, suggesting that PP2As
30 function as regulators of various signaling pathways associated with protein synthesis, cell cycle and apoptosis (Harris *et al.*, *Plant Physiol.* **121**(2): 609-617, 1999). PP2A

enzymes have been implicated as mediators of a number of plant growth and developmental processes.

In addition, PP2A enzymes play a role in pathogen invasion. In animals, a variety of viral proteins target specific PP2A enzymes to deregulate chosen cellular pathways in the host and promote viral progeny (Sontag, E., *Cell Signal* 13(1): 7-16, 2001; Garcia *et al.*, *Microbes Infect.* 2(4): 401-407, 2000). PP2A enzymes interact with many cellular and viral proteins, and these protein-protein interactions are critical to modulation of PP2A signaling (Sontag, *supra*). The proteins interacting with PP2A (*e.g.*, PP2A) can, for example, target PP2A to different subcellular compartments, or affect PP2A enzyme activity. Moreover, PP2A enzymes play a role in plants in their response to viral infection (Dunigan and Madlener, *Virology* 207(2): 460-466, 1995). Indeed, serine/threonine protein phosphatase is required for tobacco mosaic virus-mediated programmed cell death (Dunigan and Madlener, *supra*).

OsPP2A-2 (GenBank Accession No. AF134552) is a 308-amino acid subunit of a family of protein phosphatases that contains a serine/threonine protein phosphatase signature (amino acids 112 to 117).

As described above, a yeast two-hybrid approach was taken to dissect PP2A-mediated signaling events. The bait fragments used in this search and found to have interactors encode amino acids 1 to 308 and 150-308 of OsPP2A-2.

The second bait used in this Example, OsCAA90866, is a protein encoded by a complete cDNA sequence that is only known to be inducible by chilling in rice. OsCAA90866 was chosen as a bait for these interaction studies based on its relevance to abiotic stress. Investigation into the interactions involving OsCAA90866 will provide insight into the function of this poorly defined protein. The identification of rice genes involved in modulating the response of the plant to an environmental challenge, thus conferring it a selective advantage, would facilitate the generation and yield of crops resistant to abiotic stress.

Results

OsPP2A-2 was found to interact with rice putative proline-rich protein, which is possibly a transcriptional regulator, and with the seed storage protein glutelin. The search also identified five novel rice proteins interacting with OsPP2A-2: a putative
5 PP2A regulatory subunit protein also similar to rice chilling-inducible protein CAA90866 (the second bait protein of this Example); an enzyme similar to phosphoribosylanthranilate transferase that is likely involved in the plant response to pathogen infection; a disulfide isomerase, with a putative role in protein folding; a voltage-dependent ion channel protein; and a DnaJ-like protein with a putative role in the
10 pathogen-induced defense response.

The second bait protein of this Example, chilling-inducible protein CAA90866 was found to interact with itself and with six proteins. One of these is the same putative PP2A regulatory subunit protein (similar to the bait protein itself) found to interact with
15 the bait OsPP2A-2 of described in this Example. This interaction links the two networks of proteins identified in thi Example (*i.e.*, links proteins associated with biotic and abiotic stress to phosphatases). The other interactors identified in this search include a 14-3-3-like protein that is induced under various abiotic stress conditions; a pyrrolidone carboxyl peptidase-like protein with a putative role in activating signaling peptides involved in the
20 plant's response to cold stress; a novel protein containing an inositol phosphate domain likely involved in regulation of signaling events associated with cold tolerance; a novel rice homolog of wheat initiation factor (iso)4f p82 subunit with a putative role in RNA decay pathways associated with stress conditions; and a novel protein similar to plants 2-dehydro-3-deoxyphosphooctonate aldolase.

25 (

The interacting proteins of the Example are listed in Table 31 and Table 32 below, followed by detailed information on each protein and a discussion of the significance of the interactions. A diagram of the interactions is provided in Figure 6. The nucleotide and amino acid sequences of the proteins of the Example are provided in
30 Figure 16.

Some of the proteins identified represent rice proteins previously uncharacterized. Based on their presumed biological function and on their ability to specifically interact with the bait proteins OsPP2A-2 or OsCAA90866, we speculate that the proteins interacting with OsPP2A-2 represent a network involved in the rice defense response to biotic stress, and those interacting with OsCAA90866 are associated with the abiotic stress response. Importantly, the interactions identified suggest that phosphatases play a role in the regulation of both biotic and abiotic stress response in rice.

Table 31. Interacting Proteins Identified for OsPP2A-2 (Serine/Threonine Protein Phosphatase PP2A-2).

- 10 The Myriad names and the TMRI names of the clones of the proteins used as baits and found as preys are given. Nucleotide/protein sequence accession numbers for the proteins of the Example (or related proteins) are shown in parentheses under the protein name. The bait and prey coordinates (Coord) are the amino acids encoded by the bait fragment(s) used in the search and by the interacting prey clone(s), respectively. The source is the library from which each prey clone was retrieved.

Myriad/ TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (Source)
BAIT PROTEIN :			
OsPP2A-2 PN20254 (AF134552- OS002763)	<i>O. sativa</i> Serine/Threonine Protein Phosphatase PP2A-2, Catalytic Subunit (AF134552, AAD22116)		
INTERACTORS			
OsAAK63900 PN23266	<i>O. sativa</i> Putative Proline-Rich Protein AAK63900 (AC084884)	1-308	122-224 (input trait)
OsORF020300-223 PN21639 (2233(2)-OS- ORF020300 novel	Hypothetical Protein ORF020300- 2233.2, Putative PP2A Regulatory Subunit, Similar to OsCAA90866 (AAD39930; 5e ⁻⁹²) (CAA90866; 5e ⁻³³)	1-308	93-387 118-388 (input trait)
OsPN23268 PN23268 novel	Novel Protein 23268, Similar to Phosphoribosylanthranilate Transferase, Chloroplast Precursor, Fragment (AAB02913.1; 5e ⁻⁹⁵)	1-308	2x 12-200 (input trait)
OsCAA33838 PN24775	<i>O. sativa</i> Glutelin CAA33838 (X15833)	150-308	5-155 (output trait)
OsPN26645 (Contig3412.fasta.Contig1 novel)	Novel Protein PN26645, Putative Protein Disulfide Isomerase-Related Protein Precursor (BAB09470.1; e ⁻²⁸)	1-308	24-164 (input trait)
OsPN24162 (Contig3453.fasta.Contig1 novel)	Novel Protein PN24162, Porin-like, Voltage-Dependent Anion Channel Protein (NP_201551; 3e ⁻⁸⁶)	150-308	28-164 (output trait)
Os011994-D16 PN20618	Hypothetical Protein 011994-D16,	150-308	99-368

(FL_R01_P028_D16OS0119 94 novel	Similar to <i>Z. mays</i> DnaJ protein (T01643; e=0)		(output trait)
------------------------------------	---	--	----------------

Table 32. Interacting Proteins Identified for OsCAA90866 (*O. sativa* Chilling-Inducible Protein CAA90866).

5 The Myriad names and the TMRI names of the clones of the proteins used as baits and found as preys are given. Nucleotide/protein sequence accession numbers for the proteins of the Example (or related proteins) are shown in parentheses under the protein name. The bait and prey coordinates (Coord) are the amino acids encoded by the bait fragment(s) used in the search and by the interacting prey clone(s), respectively. The source is the library from which each prey clone was retrieved.

Myriad/TMRI Gene Name	Protein Name (GenBank Accession No.)	Bait Coord	Prey Coord (Source)
BAIT PROTEIN :			
OsCAA90866 PN20311 (984756_OS015052)	<i>O. sativa</i> Chilling-Inducible Protein CAA90866 (Z54153, CAA90866)		
INTERACTORS:			
OsCAA90866 PN20311	<i>O. sativa</i> Chilling-Inducible Protein CAA90866 (Z54153, CAA90866)	100-250	1-126 (output trait)
Os008938-3209 PN20215 (3209-OS208938)	<i>O. sativa</i> Putative 14-3-3 Protein (AAK38492)	100-250	4x 53-259 (input trait)
OsAAG46136 PN23186	<i>O. sativa</i> Putative Pyrrolidone Carboxyl Peptidase (AAG46136)	100-250	2x 92-222 (input trait)
OsORF020300-223 PN21639	Hypothetical Protein ORF020300-2233.2, Putative PP2A Regulatory Subunit, Similar to OsCAA90866 (AAD39930; 5e ⁻⁹²) (CAA90866, 5e ⁻⁵³)	100-250	3x 1-206 3x 1-190 (output trait)
OsPN23045	Novel Protein PN23045	100-250	2x 240-287 (input trait)
OsPN23225	Novel Protein PN23225, Similar to <i>Triticum aestivum</i> Initiation Factor (iso)4f p82 Subunit (AAA74724; e=0)	100-250	639-792 (input trait)
OsPN29883	Novel Protein PN29883, Fragment	100-250	58-175 (output trait)

10

Two Hybrid Using OsPP2A as a Bait

The bait fragment encoding amino acids 1 to 308 of *O. sativa* Serine/Threonine Protein Phosphatase PP2A-2, Catalytic Subunit (OsPP2A-2) was found to interact with *O. sativa* (rice) putative proline-rich protein, which is possibly a transcriptional regulator.

The bait fragment (*i.e.*, aa 1-308 of OsPP2A-2) includes the serine/threonine protein phosphatase signature of OsPP2A-2. One prey clone encoding amino acids 122 to 224 of OsAAK63900 was retrieved from the input trait library. Somewhat surprisingly, this prey clone does not code for the HLH domain of OsAAK63900.

5 *O. sativa* Putative Proline-Rich Protein AAK63900 (OsAAK63900) (GenBank Accession No. AC084884) is a 224-amino acid protein that includes a putative transmembrane spanning region (amino acids 7 to 23). It also contains a gntR family signature (amino acids 10 to 34) common to a group of DNA-binding transcriptional regulation proteins in bacteria (see Buck and Guest, *Biochem. J.* **260**: 737-747, 1989; Haydon and Guest, *FEMS Microbiol. Lett.* **79**: 291-296, 1991; and Reizer *et al.*, *Mol. Microbiol.* **5**: 1081-1089, 1991. This signature includes a helix loop helix (HLH) protein dimerization domain (amino acids 5 to 20) that is often found in transcription factors (see Murre *et al.*, *Cell* **56**: 777-783, 1989; Garrel and Campuzano, *BioEssays* **13**: 493-498, 1991, Kato and Dang, *FASEB J.* **6**: 3065-3072, 1992; Krause *et al.*, *Cell* **63**: 907-919, 1990; and Riechmann *et al.*, *Nucl. Acids Res.* **22**: 749-755, 1994). However, no DNA-binding motif is detectable.

Note that analysis of the amino acid sequence of OsAAK63900 also detected an Ole e I family signature (amino acids 30 to 162) including six conserved cysteines that are involved in disulfide bonds. This signature is a conserved region found in a group of plant pollen proteins of unknown function which tend to be secreted and consist of about 145 amino acids (and thus are shorter than OsAAK63900). The first of the Ole e I family of proteins to be discovered was Ole e I (IUIS nomenclature), a constitutive protein in the olive tree *Olea europaea* pollen and a major allergen (Villalba *et al.*, *Eur. J. Biochem.* **216**(3): 863-869, 1993).

25 The bait fragment encoding amino acids 1 to 308 of OsPP2A-2 (which includes the serine/threonine protein phosphatase signature of OsPP2A-2) was also found to interact with *O. sativa* OsORF020300-223.2, a novel 418-amino acid protein which has a putative PP2A regulatory subunit, similar to OsCAA90866. Two prey clones encoding amino acids 93 to 387 and 118 to 388 of ORF020300-233 were retrieved from the input trait library, which indicates that OsORF020300-223 interacts with OsPP2A-2 through a

region within amino acids 118 to 387. OsORF020300-223 includes a possible cleavage site between amino acids 50 and 51, although it appears to have no N-terminal signal peptide. OsORF020300-223 is similar to *A. thaliana* PP2A regulatory subunit (GenBank Accession No. AAD39930.1; 44.5% amino acid sequence identity; $5e^{-91}$ expectation value). OsORF020300-223 is also similar to rice chilling-inducible protein CAA90866 (GenBank Accession No. CAA90866, 68% sequence identity; $9e^{-48}$ expectation value), a protein related to chilling tolerance in rice, with which OsORF020300-223 also interacts. CAA90866 was also used as a bait protein, and the interactions identified for it are discussed later in this Example.

A BLAST analysis comparing the nucleotide sequence of OsORF020300-223 against TMRI's GeneChip[®] Rice Genome Array sequence database (http://tmri.org/gene_exp_web/) identified probeset OS015607_ at (e^{-135} expectation value) as the closest match. Gene expression experiments indicated that this gene is induced by the fungal pathogen *M. grisea*.

The bait fragment encoding amino acids 1 to 308 of OsPP2A-2 (which includes the serine/threonine protein phosphatase signature of OsPP2A-2) was also found to interact with a novel protein (23268), an enzyme similar to phosphoribosylanthranilate transferase that is likely involved in the plant response to pathogen infection. The novel protein, which was named OsPN23268, is similar to anthranilate phosphoribosyltransferase, a chloroplast precursor. Two prey clones encoding amino acids 12 to 200 of novel protein OsPN23268 were retrieved from the input trait library.

OsPN23268 is a novel 320-amino acid protein with a possible cleavage site between amino acids 43 and 44, although there does not appear to be an N-terminal peptide sequence. Analysis of the Os23268 protein sequence detected two domains originally defined in *E. coli* thymidine phosphorylase (Walter et al., *J. Biol. Chem.* 265(23): 14016-22, 1990): the glycosyl transferase family, helical bundle domain (amino acids 1 to 61) and a glycosyl transferase family, a/b domain (amino acids 66 to 303). The latter contains a beta-sheet that is splayed open to accommodate a putative phosphate-binding site (Walter et al., *J. Biol. Chem.* 265(23): 14016-14022, 1990). Two prey clones

of OsPN23268 retrieved from the input trait library and found to interact with OsPP2A-2 included sequence encoding amino acids 12 to 200 of novel protein OsPN23268. This sequence of OsPN23268 includes the glycosyl transferase family helical bundle domain and part of the a/b domain.

5 The glycosyl transferase family includes thymidine phosphorylase and anthranilate phosphoribosyltransferase enzymes. In mammalian cells, thymidine phosphorylase is identical to the angiogenic factor, platelet-derived endothelial cell growth factor (Morita *et al.*, *Curr. Pharm. Biotechnol.* **2**(3): 257-267, 2001; Browns and Bicknell, *Biochem. J.* **334**(Pt 1): 1-8, 1998), and it also controls the effectiveness of the
10 chemotherapeutic drug capecitabine by converting it to its active form (Ackland and Peters, *Drug Resist. Updat.* **2**(4): 205-214, 1999). As its name indicates, novel protein 23268 is similar to *A. thaliana* phosphoribosylanthranilate transferase (GenBank Accession No. AAB02913.1; 56.6% identity; $5e^{-95}$), an enzyme with a role in the tryptophan biosynthetic pathway which is also found in bacteria (Edwards *et al.*, *J. Mol.*
15 *Biol.* **203**(2): 523-524, 1988). In *A. thaliana*, this tryptophan biosynthetic enzyme is synthesized as a higher-molecular-weight precursor and then imported into chloroplasts to be processed into its mature form (Zhao and Last, *J. Biol. Chem.* **270**(11): 6081-6087, 1995). The *A. thaliana* anthranilate phosphoribosyltransferase is also similar to DESCA11 (GenBank Accession No. BI534445; e^{-17}), one of the genes identified in
20 *Chenopodium amaranticolor* (a plant with broad-spectrum virus resistance) which are induced during the hypersensitive response (HR) response of the plant subsequent to infection with tobacco mosaic virus and tobacco rattle tobnavirus (Cooper, B., *Plant J.* **26**(3): 339-349, 2001).

 A BLAST analysis comparing the nucleotide sequence of OsPN23268 against
25 TMRI's GeneChip[®] Rice Genome Array sequence database identified probeset OS015603_s_ at ($3e^{-41}$ expectation value) as the closest match. Our gene expression experiments indicate that this gene is induced by the fungal pathogen *M. grisea*.

 The bait fragment of OsPP2A-2 containing amino acids 150 to 308 was also
30 found to interact with the seed storage protein glutelin CAA33838 (OsCAA33838).

Glutelin CAA33838 is the major seed storage protein in rice. Its cDNA sequence was identified by Wen *et al.*, *Nucleic Acids Res.* 17(22): 9490, 1989, and the accumulation of the protein in rice endosperm occurs between five and seven days after flowering (Udaka *et al.*, *J. Nutr. Sci. Vitaminol. (Tokyo)* 46(2): 84-90, 2000). One prey clone encoding amino acids 5 to 155 of OsCAA33838 was retrieved from the output trait library.

OsCAA33838 (GenBank Accession No. X15833) is a 499-amino acid protein that includes a cleavable signal peptide (amino acids 1 to 24), as determined by analysis of the amino acid sequence. The analysis identified an 11S plant seed storage protein domain (amino acids 1 to 469; $1e^{-243}$). The 11S plant seed storage proteins tend to be glycosylated proteins that form hexameric structures. They are composed of two peptides linked by disulfide bonds and are also members of the cupin superfamily of proteins by virtue of their two beta-barrel domains. The analysis also detected this domain but localized it to a narrower region (amino acids 302 to 324). In addition, a 7S seed storage protein, C-terminal domain (amino acids 319 to 478; $602e^{-04}$), was identified which is also found in members of the cumulin superfamily. In agreement with the evidence that OsCAA33838 is a glycosylated protein, an N-glycosylation site (amino acids 491 to 494) was identified.

A BLAST analysis comparing the nucleotide sequence of OsCAA33838 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS000688.1 _ at ($e=0$ expectation value) as the closest match. Our gene expression experiments indicate that this gene is not specifically expressed in several different tissue types and is not specifically induced by a broad range of plant stresses, herbicides and applied hormones.

The bait fragment of OsPP2A-2 was also found to interact with novel protein PN26645, a putative protein disulfide isomerase-related protein precursor (also called OsPN26645). The bait fragment used in this search encodes amino acids 1 to 308 of OsPP2A-2, which includes the serine/threonine protein phosphatase signature of OsPP2A-2. One prey clone encoding amino acids 24 to 164 of OsPN26645 was retrieved from the input trait library. OsPN26645 is a 311-amino acid protein that includes a cleavable signal peptide (amino acids 1 to 17) and a predicted transmembrane domain

(amino acids 210 to 226), as determined by analysis of the amino acid sequence. A BLAST analysis against the Genpept database revealed that OsPN26645 is similar to an *A. thaliana* protein (GenBank Accession No. BAB09470.1; 32.8% identity; e^{-28}) that is similar to the rat protein disulfide isomerase-related protein precursor (GenBank Accession No.: gi5668777, 46% identity, $1e^{-63}$). As its name indicates, disulfide isomerase catalyzes the formation of disulfide bonds. This enzyme may therefore be important for proper protein folding. In mammals, disulfide isomerase in the lumen of the endoplasmic reticulum creates disulfide bonds in secretory and cell-surface proteins, and microsomes deficient in this enzyme are unable to conduct cotranslational formation of disulphide bonds (Bulledi and Freedman, *Nature* 335(6191): 649-651, 1988). Although the activity of this enzyme is not as well characterized in plants, it is likely that it serves in a similar capacity.

A BLAST analysis comparing the nucleotide sequence of OsPN26645 against TMRI's GeneChip[®] Rice Genome Array sequence database identified probeset OS002485.1 _ at (e^{-105} expectation value) as the closest match. Gene expression experiments indicated that this gene is not specifically expressed in several different tissue types and is not specifically induced by a broad range of plant stresses, herbicides and applied hormones.

The bait fragment of OsPP2A-2 was also found to interact with novel protein PN24162 (OsPN24162), a porin-like, voltage-dependent anion channel protein. The bait fragment used in this search encodes amino acids 150 to 308 of OsPP2A-2. One prey clone encoding amino acids 28 to 164 of OsPN24162 was retrieved from the output trait library. BLAST analysis of the OsPN24162 amino acid sequence indicated that this protein is most similar to a porin-like protein from *A. thaliana* (GenBank Accession No. NP_201551; 53% amino acid sequence identity; $3e^{-86}$). OsPN24162 is also similar to a rice mitochondrial voltage-dependent anion channel (GenBank Accession #Y18104; 44% identity; $2e^{-61}$), a 274-amino acid protein encoded by a cDNA found to belong to a small multigene family in the rice genome (Roosens *et al.*, *Biochim. Biophys. Acta* 1463(2): 470-476, 2000). Expression of this gene was found to be regulated in function of the plantlets maturation and organs, and not responsive to osmotic stress (Roosens *et al.*,

supra). Mitochondrial voltage-dependent ion channels are also called mitochondrial porins by analogy with the proteins forming pores in the outer membrane of Gram-negative bacteria.

5 A BLAST analysis comparing the nucleotide sequence of OsPN24162 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS007036.1 _ at (e^{-65} expectation value) as the closest match. Our gene expression experiments indicate that this gene is not specifically expressed in several different tissue types and is not specifically induced by a broad range of plant stresses, herbicides and applied hormones.

10

The bait fragment of OsPP2A-2 was also found to interact with search a DnaJ-like protein with a putative role in the pathogen-induced defense response. The bait fragment used in this search encodes amino acids 150 to 308 of OsPP2A-2. One prey clone encoding amino acids 99 to 368 of Os011994-D16 was retrieved from the output trait library. This new protein was named 011994-D16 or, because it was identified from *O. sativa*, Os011994-D16.

15 BLAST analysis of the Os011994-D16 amino acid sequence indicated that this protein is similar to maize (*Zea mays*) DnaJ protein homolog ZMDJ1 (GenBank Accession No. T01643; 84% identity; $e=0$). In eukaryotic cells, DnaJ-like proteins regulate the chaperone (protein folding) function of Hsp70 heat-shock proteins through direct interaction of different Hsp70 and DnaJ-like protein pairs (Cyr *et al.*, *Trends Biochem. Sci.* 19(4): 176-181, 1994). Heat shock proteins (reviewed in Bierkens *et al.*, *Toxicology* 153(1-3): 61-72, 2000) are stress proteins which function as intracellular chaperones to facilitate protein folding and assembly and which are selectively expressed in plant cells in response to a range of stimuli, including heat and a variety of chemicals. As regulators of heat shock proteins, DnaJ-like proteins are thus part of the plant protective stress response.

25 A BLAST analysis comparing the nucleotide sequence of Os011994-D16 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS009139.1 _ at ($e=0$ expectation value) as the closest match. Gene expression

30

experiments indicated that expression of this gene is repressed by the plant hormone jasmonic acid.

Yeast Two-Hybrid Using *O. sativa* Chilling-Inducible Protein CAA90866
(OsCAA90866) as Bait

5 The bait protein, namely *O. sativa* chilling-inducible protein CAA90866 (OsCAA90866), is a 379-amino acid protein encoded by a complete cDNA sequence related to chilling tolerance in rice. BLAST analysis indicated that OsCAA90866 is similar to the same PP2A regulatory subunit from *A. thaliana* (GenBank Accession
10 #AAD39930; 35% amino acid sequence identity; e^{-57} expectation value) that was found similar to OsORF020300-223, interactor for the bait protein PP2A-2 (see Example III, page). A BLAST analysis comparing the nucleotide sequence of the chilling-inducible protein against TMRI's GeneChip[®] Rice Genome Array sequence database identified
15 probeset OS015052 _at ($4e^{-78}$ expectation value) as the closest match. Gene expression experiments indicated that this gene is induced by cold stress.

As described in Table 32, a bait clone encoding amino acids 100 to 250 of *O. sativa* Chilling-Inducible Protein CAA90866 (OsCAA90866) was found to interact with
20 a prey clone encoding amino acids 1 to 126 of the same protein retrieved from the output trait library.

In addition, the bait clone encoding amino acids 100 to 250 of *O. sativa* Chilling-Inducible Protein CAA90866 (OsCAA90866) was found to interact with Os008938-3209. Four prey clones encoding amino acids 53-259 of Os008938-3209 were retrieved from the input trait library. Os008938-3209 is a 260-amino acid protein that includes a
25 14-3-3 protein signature 1 (amino acids 48-60) and a 14-3-3 protein signature 2 (amino acids 220 to 260), which suggests that Os008938-3209 is a member of the 14-3-3 family. BLAST analysis indicated that the amino acid sequence of Os008938-3209 shares 100% identity with that of rice putative 14-3-3 protein (GenBank Accession No. AAK38492, $8e^{-145}$). The 14-3-3 proteins interact with regulators of cellular signaling, cell cycle
30 regulation, and apoptosis. They are thought to act as molecular scaffolds or chaperones and to regulate the cytoplasmic and nuclear localization of proteins with which they

interact by regulating their nuclear import/export Zilliacus *et al.*, *Mol. Endocrinol.* **15**(4): 501-511, 2001); reviewed by Muslin *et al.*, *Cell Signal* **12**(11-12): 703-709, 2000. Since 14-3-3 proteins participate in protein complexes within the nucleus (Imhof and Wolffe, *Biochemistry* **38**(40): 13085-13093, 1999; Zilliacus *et al.*, *supra*), cytoplasm (De Lille *et al.*, *Plant Physiol.* **126**(1): 35-38, 2001), mitochondria (De Lille *et al.*, *supra*) and chloroplast (Sehnke *et al.*, *Plant Physiol.* **122**(1): 235-242, 2000), additional information would be necessary to determine where Os008938-3209 resides within the cell. Cellular localization of this prey protein could lead to a better interpretation of the significance of its interaction with chilling-inducible protein CAA90866.

10 A BLAST analysis comparing the nucleotide sequence of the Os008938-3209 protein against TMRI's GeneChip® Rice Genome Array sequence database identified probeset OS008938_s_at (e^{-61} expectation value) as the closest match. Gene expression experiments indicated that this gene is induced by salicylic acid, ABA, BAP, BL2, and 2,4D, during cold stress, and under drought conditions.

15 In addition, the bait clone encoding amino acids 100 to 250 of *O. sativa* Chilling-Inducible Protein CAA90866 (OsCAA90866) was found to interact with OsAAG46136, a pyrrolidone carboxyl peptidase from *O. sativa*. Two prey clones encoding amino acids 92-222 of OsAAG46136 were retrieved from the input trait library. These clones include 20 the pyroglutamyl peptidase I motif of OsAAG46136.

OsAAG46136 is a 222-amino acid protein that contains a pyroglutamyl peptidase I motif (amino acids 11 to 221). This motif is found in the N-terminal regions of peptide hormones (including thyrotropin-releasing hormone and luteinizing hormone releasing hormone), and it confers protease resistance to the protein (Odagaki *et al.*, *Structure Fold Des.* **7**(4): 399-411, 1999). BLAST analysis indicated that the amino acid sequence of 25 OsAAG46136 shares 100% identity with that of rice putative pyrrolidone carboxyl peptidase (GenBank Accession No. AAG46136; $4e^{-126}$). OsAAG46136 is also similar to two unknown proteins from *A. thaliana* (GenBank Accession Nos. NP_176063, $8e^{-080}$ and AAK25976.1, e^{-076} , both not described in the literature. The similarity of 30 OsAAG46136 to pyrrolidone carboxyl peptidase gives some suggestion as to the function of this poorly defined rice protein. Pyrrolidone carboxyl peptidase (Pcps) is an enzyme

that removes an N-terminal pyroglutamyl group from some proteins. It is present in many species (reviewed by Awade *et al.*, *Proteins* 20(1): 34-51, 1994) and is a valuable tool for bacterial diagnosis (most of the literature describing this protein addresses bacterial homologs). The active site of the *Pseudomonas fluorescens* Pcps has been
 5 characterized and the nature of this site (Cys-144 and His-166 are necessary for activity) suggests that it may represent a new class of thiol aminopeptidases (Le Saux *et al.*, *J. Bacteriol.* 178(11): 3308-3313, 1996). Peptidases in this protein family are necessary for processing and activation of important bioactive peptides including amyloid precursor protein (APP), strongly implicated in Alzheimer's disease (Lefterov *et al.*, *FASEB J.*
 10 14(12): 1837-1847, 2000). Furthermore, this enzyme deaminates and thus inactivates the glycopeptide anticancer agent bleomycin (Schwartz *et al.*, *Proc. Natl. Acad. Sci. USA* 96(8): 4680-4685, 1999).

A BLAST analysis comparing the nucleotide sequence of OsAAG46136 against TMRI's GeneChip® Rice Genome Array sequence database identified probeset
 15 OS013894_s _ at (e^{-8} expectation value) as the closest match. The expectation value is too low for this probeset to be a reliable indicator of the gene expression of OsAAG46136.

The bait clone encoding amino acids 100 to 250 of *O. sativa* Chilling-Inducible
 20 Protein CAA90866 (OsCAA90866) was also found to interact with protein ORF020300-2233.2 (OsORF020300-223), having a putative PP2A regulatory subunit and being similar to OsCAA90866 (see description in Example III). Three prey clones encoding amino acids 1 to 206 and three prey clones encoding amino acids 1-190 of OsORF020300-223 were retrieved from the output trait library.

25 Additionally, the bait clone encoding amino acids 100 to 250 of *O. sativa* Chilling-Inducible Protein CAA90866 (OsCAA90866) was found to interact with protein PN23045 (OsPN23045). Two prey clones encoding amino acids 240 to 287 of OsPN23045 were retrieved from the input trait library.

OsPN23045 is a 287-amino acid protein that includes an inositol P domain (amino
 30 acids 233 to 272). This domain was identified in bovine inositol polyphosphate 1-

phosphatase protein, which is involved in signal transduction (see York *et al.*, *Biochemistry* 33(45): 13164-13171, 1994). Mikami *et al.* showed that phosphatidylinositol-4-phosphate 5-kinase (AtPIP5K11) is induced by water stress and abscisic acid (ABA) in *A. thaliana*, suggesting a link between phosphoinositide signaling cascades with water-stress responses in plants (Mikami *et al.*, *Plant J.* 15(4): 563-568, 1998). Xiong *et al.* reported that FRY1, a mutant gene in *A. thaliana* encoding an inositol polyphosphate 1-phosphatase, is a negative regulator of ABA and stress signaling in this plant (Xiong *et al.*, *Genes Dev.* 15(15): 1971-1984, 2001), providing evidence that phosphoinositols mediate ABA and stress signal transduction in plants.

10 A BLAST analysis comparing the nucleotide sequence of OsPN23045 against TMRI's GeneChip[®] Rice Genome Array sequence database identified probeset OS006742.1_ at (e=0 expectation value) as the closest match. Gene expression experiments indicated that this gene is specifically expressed in leaf and stem.

The bait clone encoding amino acids 100 to 250 of *O. sativa* Chilling-Inducible Protein CAA90866 (OsCAA90866) was also found to interact with protein PN23225, which is a novel 792-amino acid protein similar to *T. aestivum* initiation factor (iso)4f p82 subunit (p82) (GenBank Accession No. AAA74724; 69.6% amino acid sequence identity; e=0). One prey clone encoding amino acids 639 to 792 of OsPN23225 was retrieved from the input trait library. The wheat protein contains possible motifs for ATP binding, metal binding, and phosphorylation (Allen *et al.*, *J. Biol. Chem.* 267(32): 23232-23236, 1992). OsPN23225 contains an MIF4G domain (amino acids 207 to 434) named after Middle domain of eukaryotic initiation factor 4G (eIF4G), and an MA3 domain (amino acids 627 to 739) also found in eIF proteins (Ponting, C.P., *Trends Biochem. Sci.* 25(9): 423-426, 2000). These domains are found in molecules that participate in mRNA decay pathways. Although the function of the bait chilling-inducible protein CAA90866 is not well defined, it appears to be a nuclear protein and its interaction with the eIF-like protein OsPN23225 supports the notion that CAA90866 participates in the rice transcriptional machinery. The identification of the OsPN23225 prey protein likely represents the discovery of a novel rice eIF.

A BLAST analysis comparing the nucleotide sequence of OsPN23225 against TMRI's GeneChip[®] Rice Genome Array sequence database identified probeset OS003249_ at (e^{-17} expectation value) as the closest match. The expectation value is too low for this probeset to be a reliable indicator of the gene expression of OsPN23225.

5

The bait clone encoding amino acids 100 to 250 of *O. sativa* Chilling-Inducible Protein CAA90866 (OsCAA90866) was also found to interact with OsPN29883, a 340-amino acid fragment that is similar to *A. thaliana* putative 2-dehydro-3-deoxyphosphooctonate aldolase (GenBank Accession No. NP_178068; $3e^{-142}$ expectation value) and pea (*Pisum sativum*) 2-dehydro-3-deoxyphosphooctonate aldolase (Kdo8P synthase) (GenBank Accession No. O50044; $3e^{-142}$ expectation value). One prey clone encoding amino acids 58 to 175 of OsPN29883 was retrieved from the output trait library. Kdo8P synthase in pea catalyzes the biosynthesis of Kdo-8-P, a component of lipopolysaccharide of plant cell walls, with high structural and functional similarities to enterobacterial Kdo8P synthase (Brabetz *et al.*, *Planta* 212(1): 136-143, 2000).

15

Summary

The interactors identified for the OsPP2A-2 bait protein (i.e., proteins that bind to OsPP2A-2) comprise a network that is speculated to be associated with the plant defense response to pathogens. Among the five novel rice proteins identified as interactors for OsPP2A-2, Os23268 is similar to the *A. thaliana* tryptophan biosynthetic enzyme anthranilate phosphoribosyltransferase. This enzyme is encoded by a gene that is similar to the DESCA11 gene involved in resistance to virus infection (Cooper, B., *Plant J.* 26(3): 339-49, 2001). While the role of tryptophan in disease resistance is unknown, tryptophan is used in the biosynthesis of indol-3-acetic acid, a plant hormone and signaling molecule. Tryptophan may thus have a role in modulation of gene expression in plants. Moreover, the glycosyl transferase function in Os23268 may be associated with disease resistance signaling pathways or with phytoalexin cellular distribution. Phytoalexins are low-molecular-weight antimicrobial compounds that accumulate in plants as a result of infection or stress, and the rapidity of their accumulation is associated with resistance in plants to diseases caused by fungi and bacteria. Taken altogether, these

30

data suggest that anthranilate phosphoribosyltransferases plays a role in the plant response to pathogen infection. Moreover, gene expression experiments confirmed that this gene is induced by the fungal pathogen *M. grisea*. Thus, the anthranilate phosphoribosyltransferase-like novel protein Os23268 is believed to be involved in the signaling and regulation pathways that mediate the response of rice to biotic stress.

Novel protein Os011994-D16, similar to DnaJ protein, is another interactor for OsPP2A-2 with a likely role in the pathogen-induced defense response. DnaJ-like proteins are known to be regulators of heat shock proteins and are thus part of the plant protective stress response. Gene expression experiments support this notion, indicating that the gene encoding the DnaJ-like protein of this Example is repressed by jasmonic acid, a component of signaling networks that provide the specificity of plant pathogen-induced defense responses (reviewed in Nurnberger and Scheel, *Trends Plant Sci.* 6(8): 372-379, 2001).

OsPP2A-2 was also found to interact with the novel protein OsORF020300-223, which is similar to *A. thaliana* PP2A regulatory subunit and to rice chilling inducible protein CAA90866 (OsCAA90866) (the second bait protein of this Example). The similarity of OsORF020300-223 to PP2A regulatory subunit validates its interaction with the PP2A-2 catalytic subunit, this interaction being consistent with the subunit composition of PP2A enzymes (Awotunde *et al.*, *Biochim Biophys Acta* 1480(1-2): 65-76, 2000). The OsORF020300-223-OsPP2A-2 interaction suggests that OsORF020300-223 participates in signaling events that involve OsPP2A-2 enzymatic activity, and the similarity of OsORF020300-223 to rice chilling-inducible protein OsCAA90866 suggests that cold tolerance may involve one of these signaling events.

OsPP2A-2 was also found to interact with rice putative proline-rich protein OsAAK63900. Though it has no known DNA-binding motif, there are indications that OsAAK63900 may play a role as a transcriptional regulator. It has an HLH domain common to transcription factors, although this domain mediates protein dimerization only. It also has a gntR family signature common to bacterial DNA-binding transcriptional regulators, although the function of this domain is not known. The

existence of the Ole e I suggests that OsPP2-2 may dephosphorylate OsAAK69300, thus regulating its function as a pollen protein, although the lack of data on the Ole e I signature function makes this possibility more difficult to argue. Evidence also exists that PP2A proteins regulate the DNA-binding activity of transcription factors in plants
5 Vazquez-Tello *et al.*, *Mol. Gen. Genet.* **257**(2): 157-166, 1998) and mammalian cells (Wadzinski *et al.*, *Mol. Cell Biol.* **13**(5): 2822-2834, 1993). Therefore, it is most likely that the OsPP2A-2-OsAAK63900 interaction occurs in the nucleus and that it plays a role in regulating transcriptional events in rice.

10 Other proteins found to interact with OsPP2A-2 include a disulfide isomerase with a putative role in protein folding (novel protein OsPN26645), a voltage-dependent ion channel protein (novel protein OsPN24162) and the seed storage protein glutelin (OsCAA33838). The biological significance of these interactions is unclear. Analysis of
15 II phosphorylation sites. It is possible that the phosphorylation state of glutelin determines its function or stability, and its interaction with OsPP2A-2 may occur during dephosphorylation of glutelin. Alternatively, this interaction may result in localization of OsPP2A-2 and thereby affect events downstream of OsPP2A-2-dependent
20 dephosphorylation. Given the presence of a disulfide bond between the two peptide chains of typical plant seed storage proteins, it is interesting that OsPP2A-2 also interacts with a putative protein disulfide isomerase (OsPN26645). Perhaps OsPP2A-2 interacts with other enzymes to create a co-translational modification complex. Additional yeast-two-hybrid data may clarify the purpose of these interactions. However, given the
25 association of PP2A with other proteins involved in biotic stress responses, the aforementioned associations could also be involved in biotic stress responses.

The chilling-inducible protein CAA90866 was found to interact with itself and with six proteins. These proteins are speculated to interact as components of a network of proteins relevant to the rice response to cold stress. This hypothesis finds support in
30 gene expression experiments, which confirmed that the gene encoding the chilling-inducible protein is induced by cold. One of the interactors is the putative 14-3-3 protein

Os008938-3209. The relationship to chilling tolerance of the bait protein OsCAA90866 suggests that its interaction with Os008938-3209 may be associated with cold tolerance. Gene expression experiments showed that this protein is induced under a broad range of stress conditions. Its activation probably allows its interaction with a number of stress proteins. Given the function of 14-3-3 proteins as molecular chaperones, Os008938-3209 may act as a molecular glue for these interactions to preserve protein complex stability in membranes, or it may coordinate interactions involving transcription factors associated with stress genes. Subcellular localization of Os008938-3209 may further clarify the significance of its interaction with OsCAA90866.

Another interactor for OsCAA90866 is a pyrrolidone carboxyl peptidase-like protein (OsAAG46136). The putative pyrrolidone carboxyl peptidase function of OsAAG46136 suggests that it participates in processing and/or activation of substrate proteins, and these proteins may be important to the plant response to chilling. Peptidase activity has been associated with regulation of signaling. Carboxypeptidases, for instance, hydrolytically remove the pyroglutamyl group from peptide hormones, thereby activating these signaling molecules. A carboxypeptidase regulates Brassinosteroid-insensitive 1 (BRI1) signaling in *A. thaliana* by proteolytic processing of a protein (Li *et al.*, *Proc. Natl. Acad. Sci. USA* **98**(10): 5916-5921, 2001). Based on its ability to interact with chilling-inducible protein and on the role of the latter in chilling tolerance, it is speculated that the carboxypeptidase-like protein OsAAG46136 may have a role in activating signaling molecules/hormonal peptides that are involved in the plant response to cold stress.

The interactions of OsCAA90866 with OsPN23045, a protein with a putative inositol phosphate function, and with OsPN23225, a rice homolog of wheat initiation factor (iso)4f p82 subunit, provide further insight into the function of the bait protein. Phosphoinositols are known to mediate ABA and stress signal transduction in plants (Mikami *et al.*, *Plant J.* **15**(4): 563-568, 1998; Xiong *et al.*, *Genes Dev.* **15**(15): 1971-1984, 2001). The putative inositol phosphatase protein OsPN23045 may function in a similar way and its interaction with the chilling-inducible protein may be associated with regulation of cell signaling events that relate to cold tolerance. The prey protein

OsPN23225 likely represents a novel rice eIF. The eIF proteins have a role in RNA processing pathways (Ponting C.P., *Trends Biochem. Sci.* 25(9): 423-426, 2000) and stress is typically associated with an abundance of RNA transcripts. Based on this information and on the relationship that CAA90866 has to chilling tolerance, the
5 OsCA90866- PN23225 interaction is speculated to control translational events related to cold stress.

Finally, OsCAA90866 interacts with and is similar to the same putative PP2A regulatory subunit protein OsORF020300-223 found to interact with the bait protein
10 OsPP2A-2. This interaction provides a link between the two networks of this Example and suggests the involvement of OsPP2A-2 in both biotic and abiotic stress response pathways (see diagram in Appendix 1). Based on the observed interactions and on sequence similarities among the proteins involved in these interactions, OsPP2A-2 appears to regulate both biotic and abiotic stress response pathways. Thus, the two
15 pathways, though independent, are speculated to be linked through protein phosphatases, and that these enzymes likely mediate the plant's stress response by dephosphorylation of the proteins participating in these pathways. In this scenario, it is possible that the self-interaction observed for OsCAA90866 participates in the creation of multicomponent phosphatase complexes. Furthermore, the interaction of OsCA90866 with the aldolase-
20 like protein OsPN29883 suggests that the aldolase needs to be dephosphorylated for activation/inactivation, and that this novel protein may have roles during stress responses based upon the other interactions and the gene expression patterns of the chilling-inducible protein.

Moreover, OsORF020300-223 the *A. thaliana* regulatory A subunit of protein phosphatase 2A (PP2A-A) has been implicated in the regulation of auxin transport in *A. thaliana* (Garbers *et al.*, *EMBO J.* 15(9): 2115-2124, 1996). The phytohormone auxin controls processes such as cell elongation, root hair development and root branching. Since OsORF020300-223 is also similar to and interacts with chilling-inducible protein
30 CAA90866, it is possible that the latter may be involved in auxin transport.

EQUIVALENTS

Those skilled in the art will recognize, or be able to ascertain, using no more than.
routine experimentation, numerous equivalents to the specific embodiments described
specifically herein. Such equivalents are intended to be encompassed in the scope of the
5 following claims.

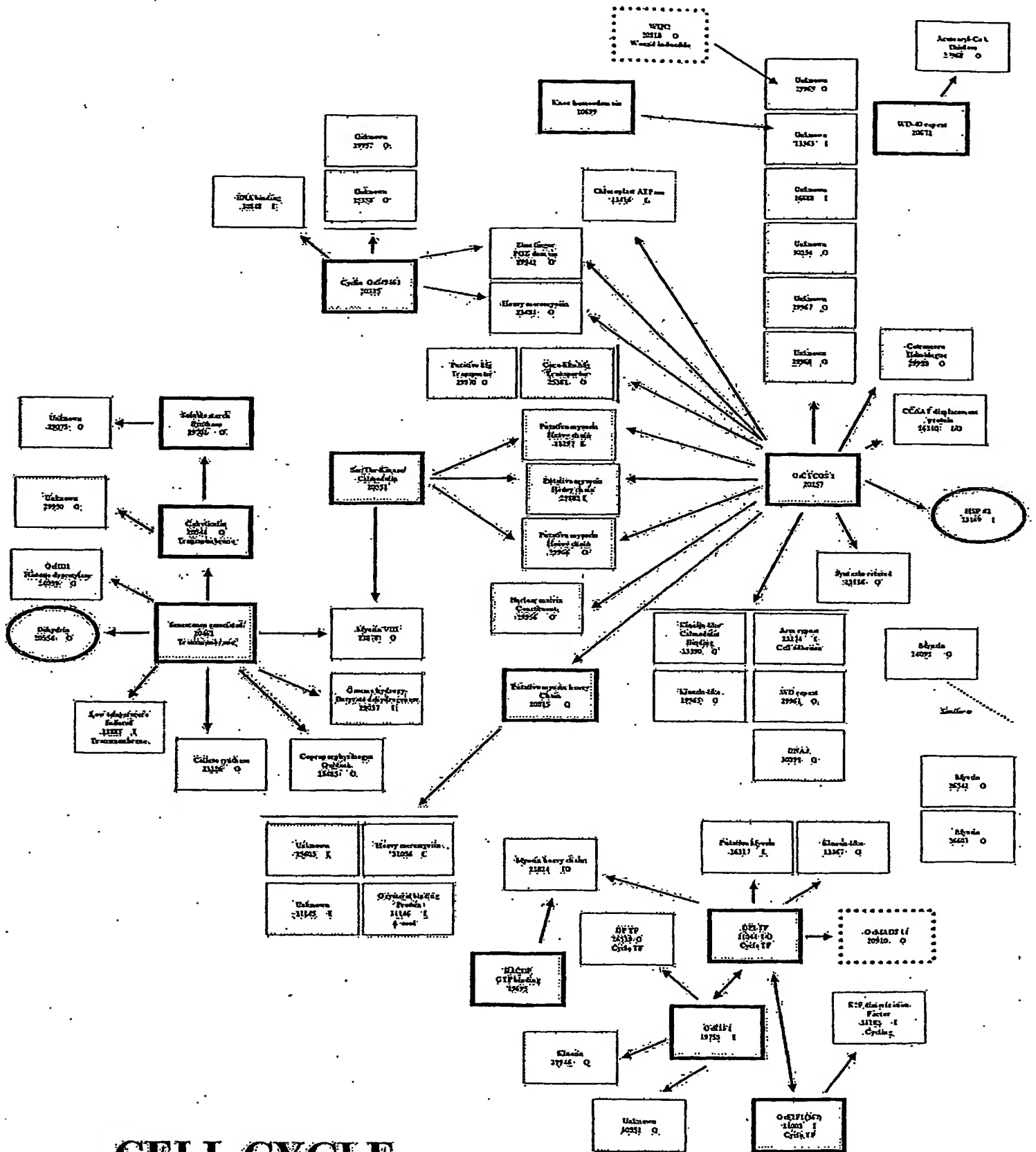
Claims

1. An isolated nucleic acid molecule encoding a cell proliferation related polypeptide, wherein the polypeptide binds to a fragment of a protein selected from the group consisting of OsE2F1, Os018989-4003, OsE2F2, OsS49462, OsCYCOS2, OsMADS45, OsRAP1B, OsMADS6, OsFDRMADS8, OsMADS3, OsMADS5, OsMADS15, OsHOS59, OsGF14-c, OsDAD1, Os006819-2510, OsCRTC, OsSGT1, OsERP, OsCHIB1, OsCS, OsPP2A-2, and OsCAA90866.

CELL PROLIFERATION-RELATED POLYPEPTIDES AND USES THEREFOR**Abstract of the Disclosure**

Disclosed are proteins, and nucleic acids encoding such proteins, involved in or associated with cell proliferation, senescence, differentiation, development, and stress response in plants. Also disclosed are uses for such proteins.

Fig. 1 (Left)
(1 of 3)



CELL CYCLE

Fig. 1 (Right)
(2 of 3)

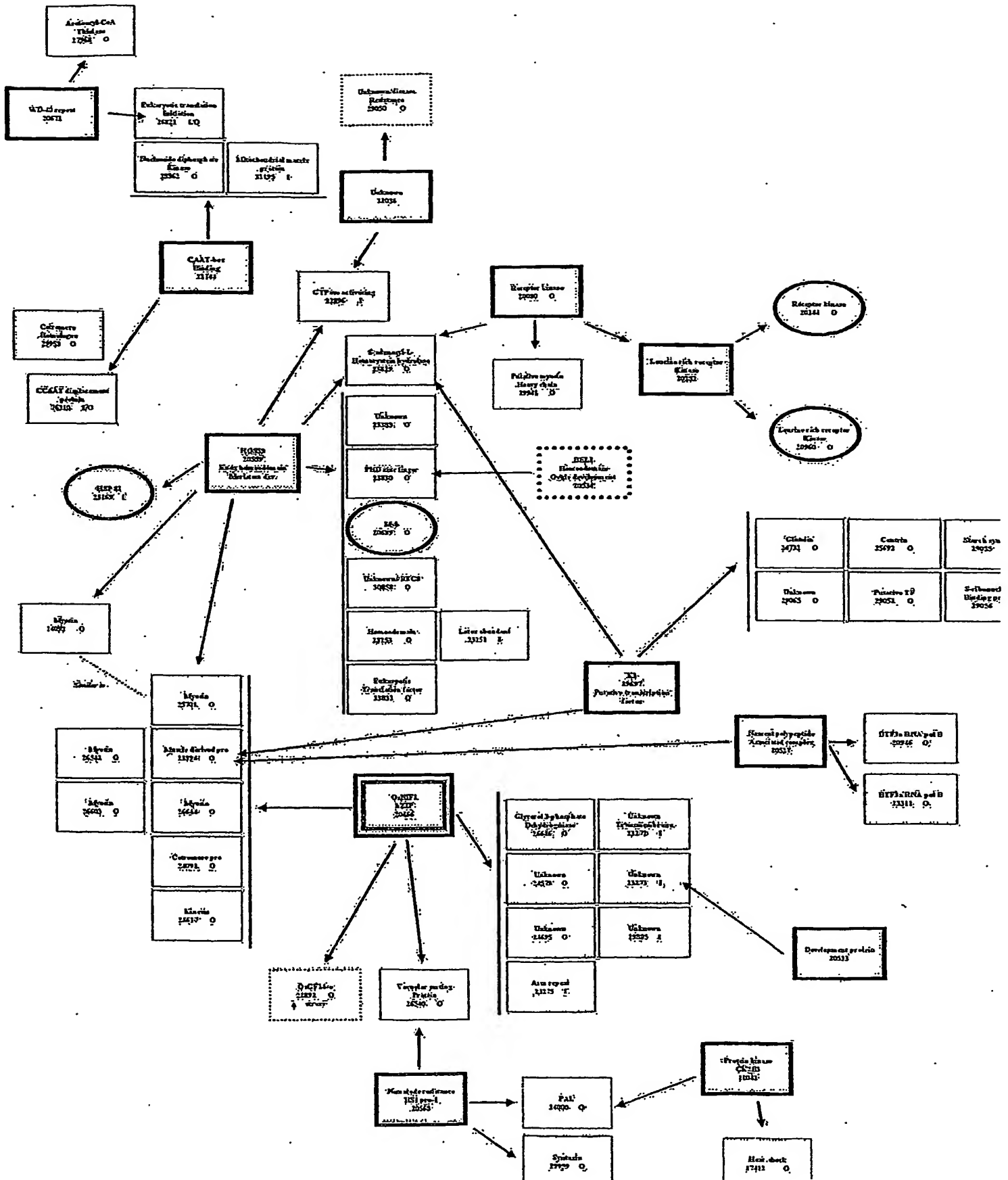
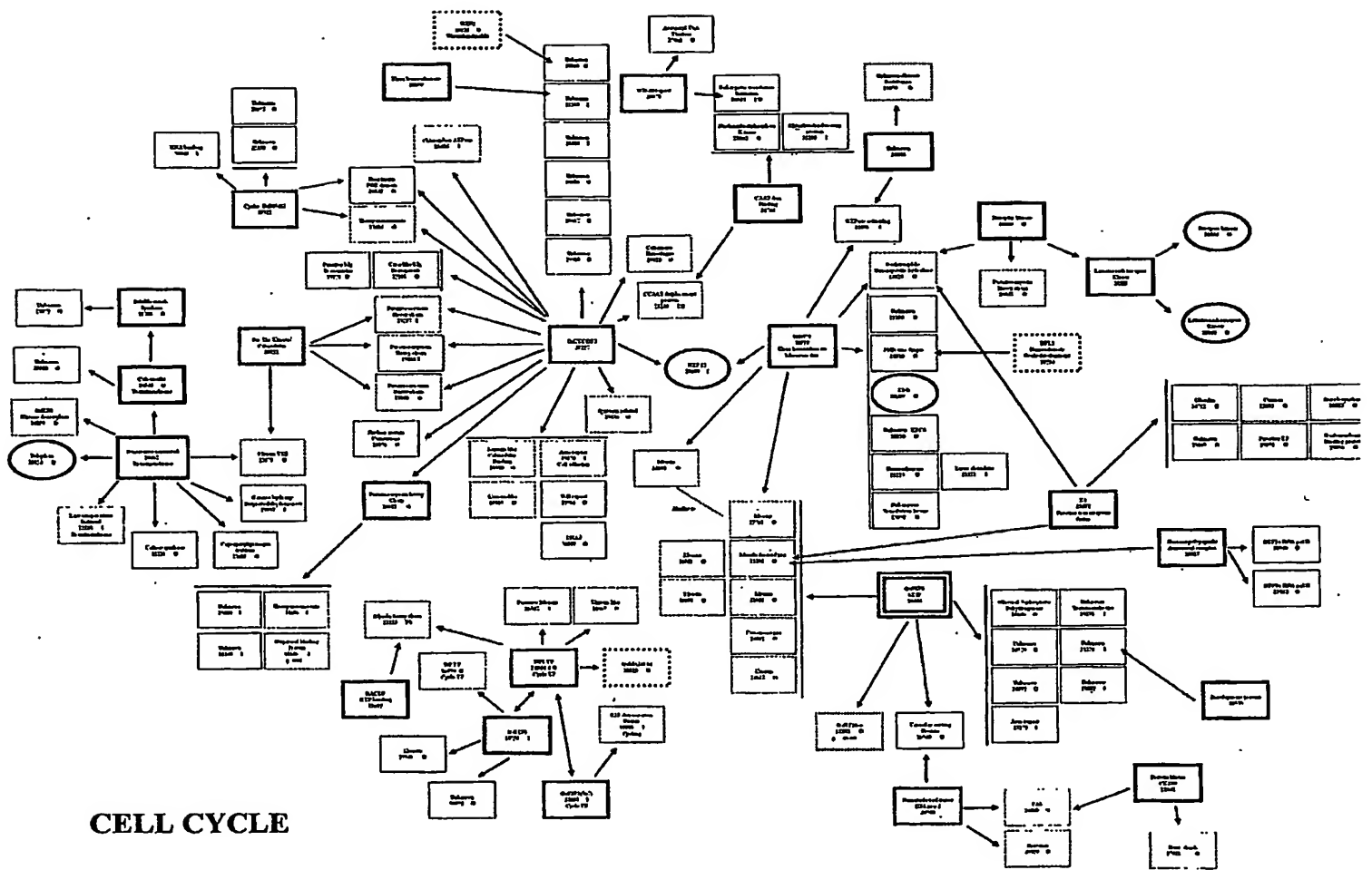


Fig. 1
(3 of 3)



APPENDIX 1

INTERACTIONS FOR OsMADS45, OsRAP1B, OsMADS6, OsFDRMADS8, OsMADS1, OsMADS3, OsMADS5,

The protein names and Myriad short gene name
 I = Input Library; O = Output Library

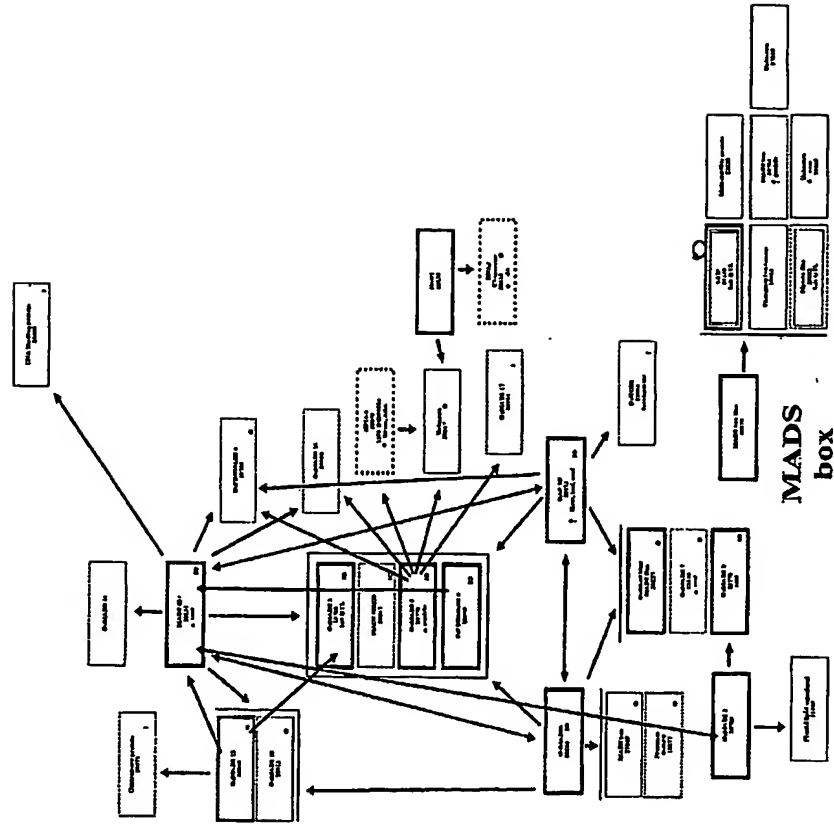


Fig. 2

APPENDIX 3

FIGURE 1. Clustal amino acid alignment of the MADS Box proteins described in the invention disclosure: Os008339, OsFDRMADS6, OsFDRMADS8, OsMADS1, OsMADS3, OsMADS5, OsMADS6, OsMADS7, OsMADS8, OsMADS13, OsMADS14, OsMADS15, OsMADS17, OsMADS18, OsBAA81880, OsMADS45, OsRAP1B, OsMADS6, and OsPN29949. Based on amino acid coordinates of the known domains in MADS6 (Moon et al., 1999), four regions, MADS Box, I region, K domain and C-terminal region, are mapped for all proteins. Vertical colored boxes (red being most similar, then orange, green, light blue, and the weakly similar blue) above the amino acid alignments reflect degrees of similarity for each region.

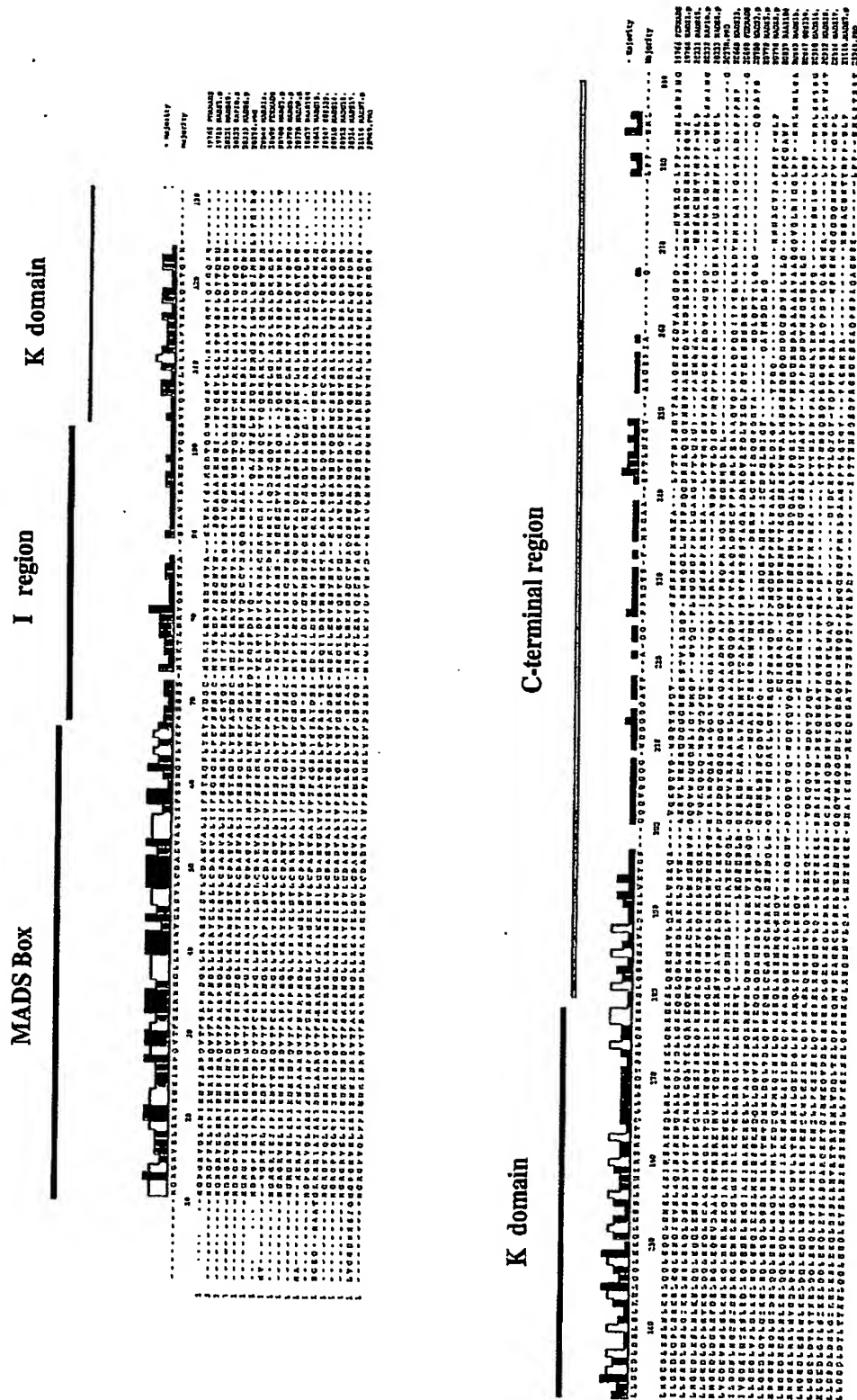


Fig. 3A

PRO70122

FIGURE 2. Phylogentic tree based on the clustal amino acid alignment of the rice MADS Box proteins.

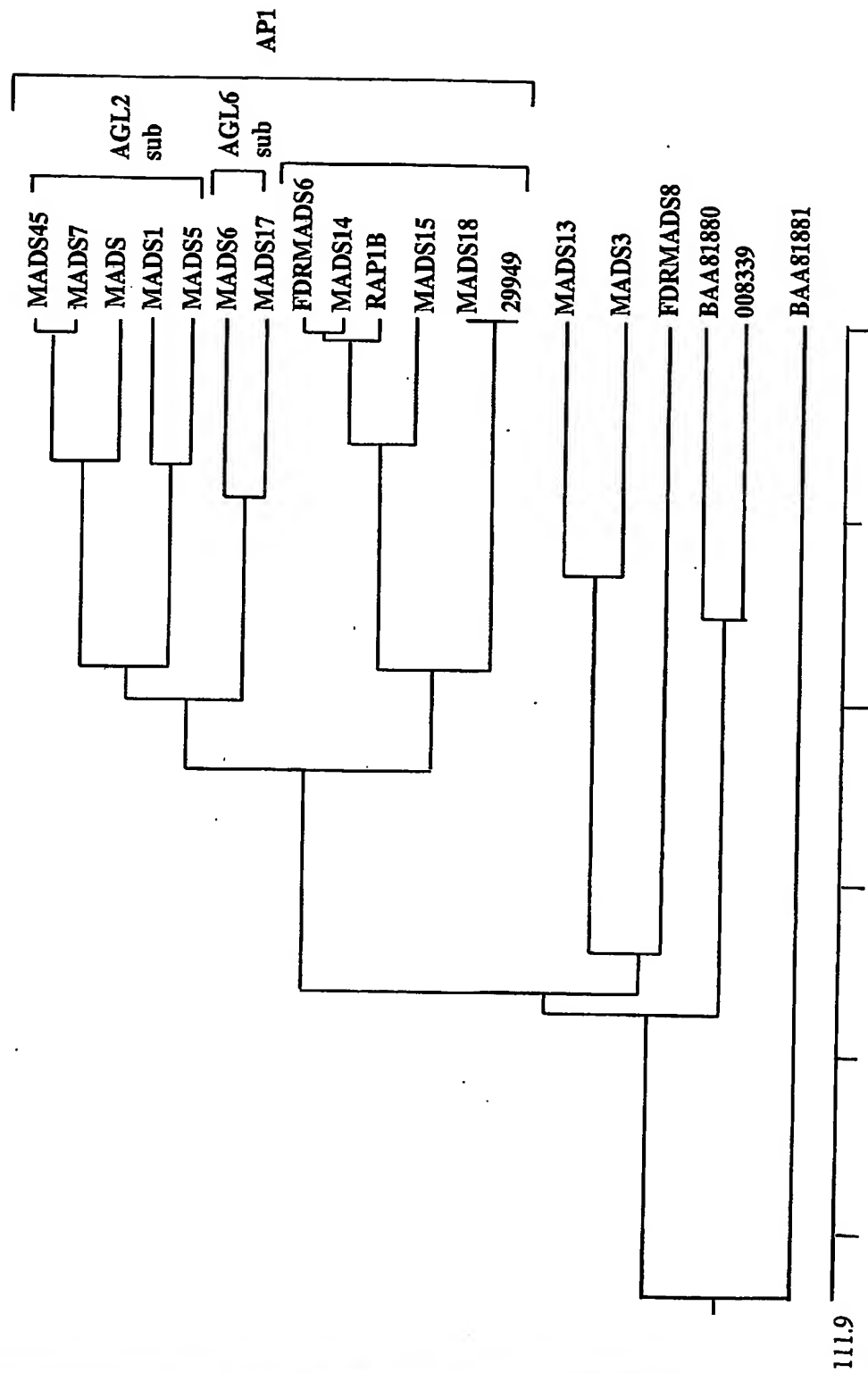


Fig. 3B

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

Fig. 4

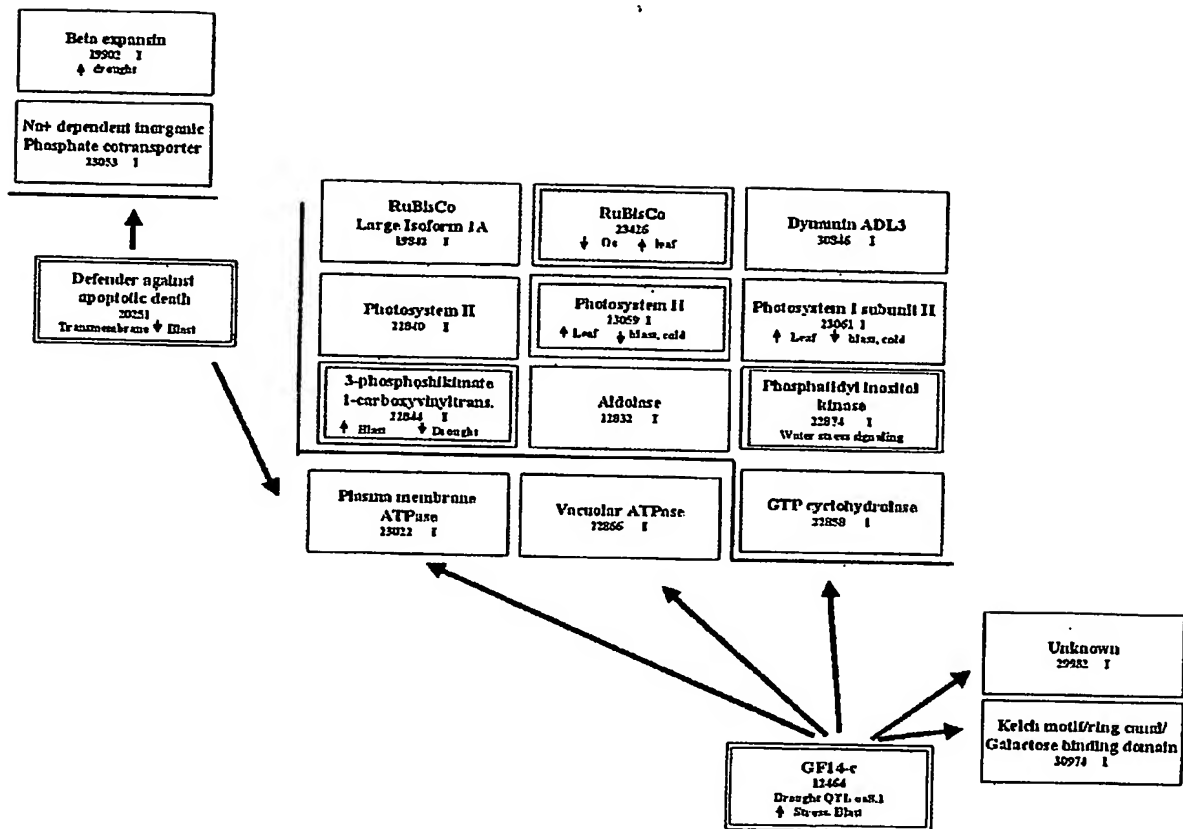
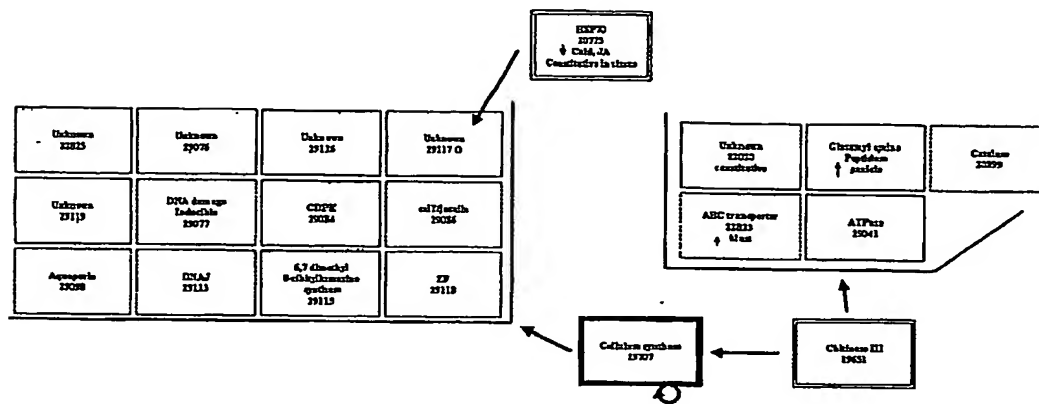


Fig. 5

CHITINASE AND CELLULOSE SYNTHASE INTERACTIONS



APPENDIX 1

INTERACTIONS FOR OsPP2A-2 AND OsCAA90866

The protein names, Myriad short gene names (into parentheses), and TMRI gene names, when available, are given.
I = Input Library; O = Output Library

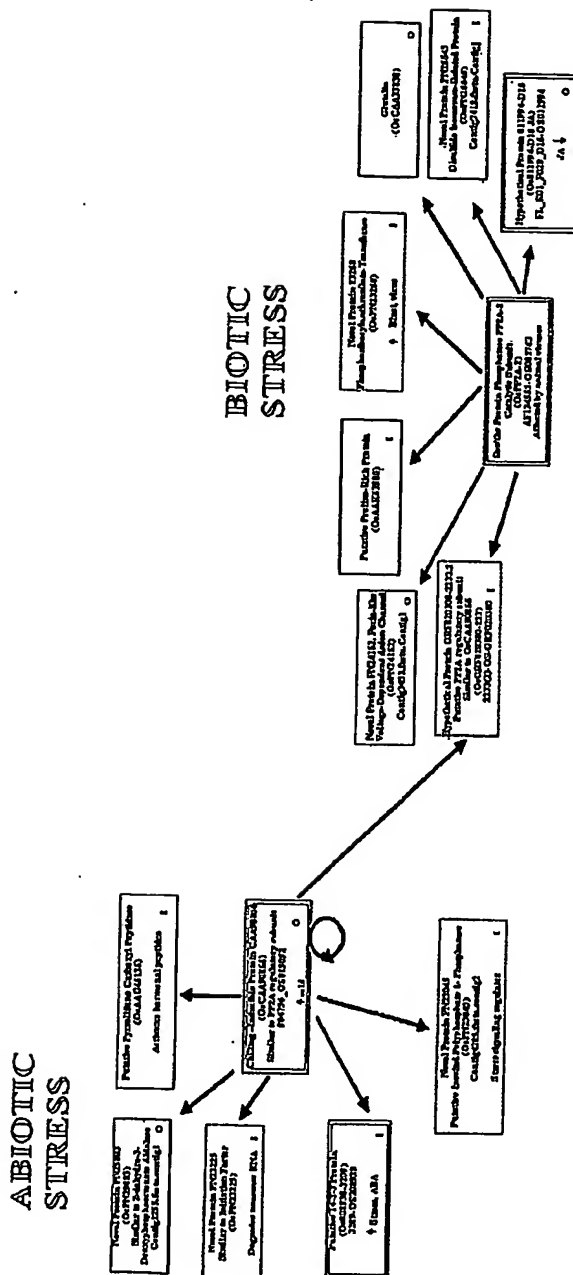


Fig. 6

AILKDISERAQKGGSFSYSRFFAVGLFRLLLEANATEPTILDKLCAALNINKRSVDRDLV
YRNILSKLVQAKELLKEYVEREKKKREERSETPKSNEAVTKFDGSLNSMRH*

>23186 OsAAG46136

MGSEGPSGVTVHVTGFKKFHGV AENPTEKIVRNLESFMEKRGLPKGLTLGSCTVLETAG
QGGGLPLYEVFESAIVDKEYGLNDQGGQVILLHFGVNSGTTRFALENQAIN EATFRCPDEL
GWKPQRAPIVSSDGSISNLRKTTVPVNEVNKSLQQMGFDVAPSDDAGR FVCNYVYYQS
LRF AEQRGIKSLFVHFPLFTTISEEVQMN FVATLLEVLASQNYAQ*

>23225 OsPN23225

MEKDHQPVISLRPGGGGGGPRPGR LFS PAF AAAASGSGDLLRSHVGGASKIGDPNF EVR
ERVRYTRDQLELREIVDIPEAILKQEIDIELHGEDQIWGRPESDVQVQTQTQAQPHNRY
GETDNRDWRARTVQPPAANEEKSWDNIREAKAAHASSGRQQEQVNRQDQLNHQFASK
AQVGPTPALIKAEVPWSARRGNLSEKDRVLKTVKGILNKLTP EKFDLLKGQLMESGITT
ADILKDVISLIFEKAVFEPTFCPMYAQLCSDLNEKLPSFPSEEPGGKEITFKRVLLNNCQE
AFEGAESLRAEIAKLTGPDQEMERRDKERIVKLRTLGNIRLIGELLKQKMVPEKIVHHIV
QELLSGSPDKKACPEEENVEAICQFFNTIGKQLDENPKSRRINDTYFIQMKELTTNLQLA
PRLRFMVRDVVDLRSNNWVPRREEIKAKTISEIHDEAIKTLGLRPGATGLTRNGRNAPG
GPLSPGGFPMNRP GTGGMMPGMPGTPGMPGSRKMPGMPGLDNDNWEVPRSKSMPRG
DSL RNQGPLLNKPSSINKPSSINSRLLPHGSGALIGKSALLGSGGPPSRPSSLMASPTHTPA
QTAPSPKPVSAAPAVVPVTDKAAGSSHEMPAAVQKKT VSLLEEYFGIRILDEAQQCIEEL
QCPEYYSEIVKEAINLALDKGPNFIDPLVRLLEHLHTKKIFKTEDLKTGCLLYAALLEDIG
IDLPLAPALFGEVVARLSLSCSLSEFVVEEILKAVEDTYFRKGIFDAVMKTMGGNSSGQA
ILSSHAVVIDACNKLLK*

>23266 OsAAK63900

MAGAPRGLVLLGVCAVLM AVAVGGEAASVVVGTA KCADCTRKNMKAEDAFKNLQV
AIKCKNGNGEYESKATGKLDGTGAFSVPLDADLHSSDCIAQLHSATNEPCPGQEPSKIVP
MSEGTFAAVAGKTHYRSALCASVTICGPIKKKIJDHFHKKPVPPKPEPKPEPPKPKPEPEH
PFLDHIHKKEKHFFDHFHKKPVPPKPEPKPEPKQPAPEYHNPSPPAN*

>23268 OsPN23268

VLETLIGGDHFSEEEAEATLRL LLEEENEARIAAFLVLLRAKGETYEEIVGLAKAMIGCC
VRVDGLDDAVDIVGTGGDGADTVNISTGSTILAAAAGAKVAKQGSRASSSACGSADGI
KRCVNEVGVGFMMSANYHPAMKIVKPVRRKKLKIKT VFNILGPLLNP ARVPYAVIGVYH
ENIVTKMAKAAQKFGMKRALVVHSGKLDEISPLGPGYILDVTPRKIEKMLFDPLDFGIPR
CTLEDLKGGDPAFNAKVLQDVL AGQRGSIADALVLNAAASLLVSGKVNTLHDGVALA
QETQRSGEAINTLESWIKISNVSTSDN*

>24162 OsPN24162

MAMKGPGLFSDIGKRAKDLLTKDYTYDQKLT VSTVSSSGVGLTSTAVKKGGLYTLDVS
SVYKYKSTLV DVKVDTESNISTTLTVFDVLPSTKLVT SVKL PDYNSGKVEMQYFHENAS
FATAVGMKPSPVVEFSGTAGAQGLAFGAEGFD TATGKFTKYSAAGVTKPDYHAAIV
LADKGDTVKVSGVYHLDDKQKSSVVAELTRRLSTNENTLT VGGLYKVDPETAVKARL
NNTGKLAALLQHEVKPKSVLTISGEFDTKALDRPPKFLALALRP*

>24775 OsCAA33838

MASSVFSRFSIYFCVLLLCHGSMAQLFNPSTNPWHSPRQGSFRECRFDRLQAFEPLRKVR
SEAGVTEYFDEKNELFQCTGTFVIRRVIQPQGLLVPRYTNIPGVVYIIQGRGSMGLTFPGC
PATYQQQFQQFSSQGSQSQKFRDEHQKIHQFRQGDIVALPAGVAHWFYNDGDAPIVA
VYVYDVNNNANQLEPRQKEFLLAGNNNRAQQQQVYGSSIEQHSGQNIFSGFGVEMLSE
ALGINAVAAKRLQSQNDQRGEIHHVKNGLQLLKPTLTQQQEQAQAQDQYQQVQYSERQ

Figure 7
Page 2 of 32

QTSSRWNGLEENFCTIKVRVNIENPSRADSYNPRAGRITSVNSQKFPILNLIQMSATRVNL
YQNAILSPFWNVNAHSLVYMIQGRSRVQVVSNGFKTVFDGVLVPAQLLIIPQHYAVLKK
AEREGCQYIAIKTNANAFVSHLAGKNSVFRALPVDVVANAYRISREQARSLKNNRGEEH
GAFTPRFQQQYYPGLSNESESE*SE*

>26645 OsPN26645

MATRLLCWTALLPIAATAAASPLPEACPVPTAAEEILGPGGTCTTLDRRGDPVGVIEG
DEVTLAKAITLLHMNKDDYIAVLFYASWCPFSQECKPNFEILASLFPSIRHFAFEESSIRPS
IISRYGIHGFPITLFLNSTMRVR YHGPRTVKSLAAFYRDVSGFDVSMTSEAVLHSVDGIE
LKKDAEQENCPCFWWARSPEKILQQDTYLALATAFVILRLLYLLFPKIGSFAKRAWRRHT
LFPNLVGVHEYFFTYLEQARHKFFRLYPSKRGNLQEGARNATAWASKSLASVSIGEPSTI
GRTNSTNELR*

>29883 OsPN29883

PRVRCRLCSSSSCSISSISHSPISIPSSALPRRSSPCVGLGAAAASEAMDASSVALYGQLK
AAQPFLLAGPNVIESEEHVLKMAKHIGITTKLGLPLVFKSSFDKANRTSSKSFRGPGL
EGLKILEKVKATYDIPVVDVHESHQCEAAGR VADIIQIPAFFCRQTDLLVAAAKTGKIN
IKKGQFCAPSVMANSAEKIRLAGNQNMVCERGTMTFGYNLIVDPNFEWLREANCPV
VADVTHALQQPAGKKLDGGGVASGGLRELIPCIARTSAVGVGDGIFMEVHDDPLNAPC
DGPTQWPLRNLEELLEELIAIARVTKGKKPLKIDLTPEKE

>19651 OsCHIB1

MVNGYLFREYIGAQFTGVRFSQVVPVNPGLSFHFI LAFAIDYFMATQSSKPAPANGVFAPY
WDTANLSPA AVAAKAHPNLSVILALGGDTVQNTGVNATFAPTSSVDAWVRNAADS
VSGLDAYGLDGDVDVYEHFAAGVDTFVECIGRLLTELKARHPNIATSIAPFEHPVVQRY
YQPLWRRYAGVIDYVNFQFYGYGANTDVATYVMFYDEQAANYPGSKLLASFKTGNVT
GLLSPEQGIAGAKELQRQGLPLFIWSADSSMVSSYKFEYETKAQEIVANH*

>19707 OsCS

MAANAGMVAGSRNRNEFVMIRPDGDAPPPAKPGKSVNGQVCQICGDTVGVVSATGDVF
VACNECAFPVCRPCYEYERKEGNQCCPQCKTRYKRHKGS PRVQGDEEEEDVDDLNEF
NYKHGNGKGPEWQIQRQGEDVDLSSSSRHEQHRIPLTSGQQISGEIPDASPRHSIRSGT
SSYVDPSVPVPVRIVDPSKDLNSYGINSVDWQERVASWRNKQDKNMMQVANKYPEAR
GGDMEGTGSNGEDIQMVDARLPLSRVPIPSNQLNLYRIVILRLILMFFFQYRVTHPVR
DAYGLWLVSVICEIWFALSWLLDQFPKWYPINRETYLDRLALRYDREGEPSQLAPIDVF
VSTVDPLKEPPLITANTVLSILA VDYVPDKVSCYVSDDGSAMLTFEALSETAEFARKWV
PFCKKHNI EPRAPEFYFAQKIDYLDKDIQPSFVKERRAMKREYEEFKVRINALVAKAQK
VPEEGWTMADGTAWPGNNPRDHPGMIQVFLGHSGGLDTDGNELPRLVYVSREKRPGF
QHHKKAGAMNALIRVSAVLTNGAYLLNVDCDHYFNNSKALREAMCFMMDPALGR

>20775 OsHSP70

MAGKGEGPAIGIDLGTTYSCVGVWQHDRVEIANDQGNRTTPSYVGFTDSERLIGDAAK
NQVAMNPINTVFD AKRLIGRRFSDASVQSDIKLWPFKVIAGPGDKPMIVVQYKGEKQF
AAEEISSMVLIKMREIAEAYLGTTIKNAVVTVPAYFNDSQRQATKDAGVIAGLNVMRIIN
EPTAAAIAYGLDKKATSVGEKNVLIFDLGGGTDFVSLLTIEEGIFEVKATAGDTHLGGED
FDNRMVNHVQEFKRNKKDITGNPRALRLRTACERAKRTLSSTAQTIEIDSLYEGID
FYSTITRARFEELNMDLFRKCMPEVEKCLRD AKMDKSSVHDVVLVGGSTRIPRVQQLLQ
DFFNGKELCKNINPDEAVAYGA AVQAAILSGEGNEKVQDLLLLDVTPLSLGLTAGGV
MTVLIPRNTTIPTKKEQVFSTYSDNQPGVLIQVYEGERTTRDNNLLGKFELSGIPPAPRG
VPQITVCFDIDANGILNVSAEDKTTGQKNKTTITNDKGRLSKEEIEKMVQGAEKYKSEDE

EHKKKVESKNALENYAYNMRNTIKDEKIASKLPAADKKKIEDAIDQAIQWLDGNQLAE
 ADEFDDKMKELEGICNPIAKMYQGAGADMAGGMEDEDDAPPAGGSGAGPKIEEVD*
 >20899 OsCATA

MDPCKFRPSSSFDTKTTTTNAGAPVWWDNEALTVGPRGPILLEDYHLIEKVAHFARERIP
 ERVVHARGASAKGFFECTHDVTDITCADFLRSPGAQTPVIVRFSTVIHERGSPETIRDPRG
 FAVKFYTREGNWDLGNNFPVFFIRDGIKFPDVIHAFKPNPRSHVQEYWRVFDLFLSHHPE
 SLHTFFFLFDDVGIPTDYRHMDGFGVNTYTFVTRDAKARYVKFHWKPTCGVSCLMDDE
 ATLVGKKNHSHATQDLYDSIAAGNFPWKLQVQVIDPEEEERFDPLDDTKTWPEDEV
 PLRPVGRVLNLRNVNFFNENEQLAFGPGLVVPGIYSSDDKMLQCRVFAYADTQRYRL
 GPNYLMPLVNPAPKCAHHNNHYDGAMNFMHRDEEVDYPSRHAPLRHAPPTPITPRPVV
 GRRQKATHKQNDFKQPGERYRSWAPDRQERFIPLRRRVAHPKVSPELRAIWNVYLSQC
 DESLGVKIANRLNVKPSM*

>22020 OsPN22020

PRVRTVASARDLKNVNWNRNGDLKPYAVVWIDDGAKCSTRVDLDNADNPTWDDKLT
 PLPPSTRLLDAVLVLDVVHANATDGVKPLVGSARLPLRDVLADTGIGARASRSLRLKRP
 SGRPHGRLEVRVAVREPKEYYDPSAPYAPYHQQTNRDPYAYGNTTTGGYGYAYGGT
 PFL*

>22154 OsPN22154

MAAAEQSAEQFRGQARLPGFAAPRRYDLRLVPDLGCAFTGSVDVSVDVTAPTRFLVL
 NAAELEVSPGGVQFKPHGAEQELHPAEVTNVPEDILIRFNEVLVPGEGTLVIAFKGTLN
 DKMHGFYRSVYELNGEKKNMAVTQFEPADARRCFPCWDEPSFKAIFKITLEVPSETVAL
 SNMPVVEEKVNGLIKAVYFQETPIMSTYLAVIVGMFDYVEAFTTDGTRVRVYTQVGK
 SAQGKFALEVAVKTLVLFKEYFAVPYPLPKMDMIAIPDFASGAMENYGLVITYRETALLF
 DEKHSAAANKQORVAVVVAHELAHQWFGNLVTMEWWTHLWLNNEGFWVSYLAADN
 FFPEWNVWTQFLEESTTGFKLDALAGSHPIEVDVNHVDEIDEIFDAISYRKGAAVIRMLQ
 SYLGAETFQKSLAAAYIEKFAYSNAKTEDLWAALEEGSGEPVKTLMHSWTKQQGYPPVN
 VKLKDGGKLEMEQTQFLSSGAEGVGQWVVPITLCCCSYSRQEKFLFNGKQEDFNLSGLV
 ECQKKEDFWIKLVNVTGTFYRVSIDEELASRLRYAIEANKLSAADRYGVLDPTYALCM
 AGKQKLVSLLHLIAAYKDETEYTVLATSLSIVEMVAVAAPEGLGKLKFLIDFLEPPAQR
 IGWDAKSGEGHLDALLRGTLTALAEKGHEATINEAVRRFNIFVEDRETPLLPPDVRKAA
 YVALMQTVNKSNRAGYESLLKIYKETDLSQEKVRILGSLASCPDPDVVRDTLDFMLSPE
 VRNQDSIFLLRGVGAAGHEVAWTWLKEKWYISDTFSGTLLTYFVSTTVSPLRTDEM
 GDDAEFFKSRRTKANIARTVKQSIERVRINAKWVESTRAEANLGNVLKEISHDH*
 >22823 OsPN22823

MDESGRGTGDDHGRETKDAAAAASSSSGKKVPLFSLFRYADRLDVLLMVVGTVGALG
 NGISQPLMTVLFNGVINSFGANTSGSVLRSVTKVKVSCWTMAGERQSARIRSLYLKAVL
 RQDITFFDTEMTTGEAVSRMSSDTLLIQGALGEKGGKLVLLSSFFIGGFIAFTRGWLLTL
 VMLTSLPLIAIASAVSAQALTRVSSKRQTSYSDAGDTVEQTIGSIRTVVSFNGEKKAIAM
 YRNFIKKSYPATIEEGITGFGMGSMCVVFGSYGLAFWYGGKLIIEKGYTGKIMTILF
 AVLTGNATPAVAAVVEGQSAAYNLFKTIERKPEIDSDDNNGMVLEDNMNGDIELKDVFYF
 RYPARPEQLILDGLSLQVASGTTMAIVGESGSGKSTVISLVERFYDPQSGEVLIDGISIKKL
 RLDWIRGKIGLVSQEPLLFMAIKDNITYGKKDATLEEIKRAAELANANFIDKLPNGYD
 TLVGQRGTQLSGGQKQRIAIARAILKDPKILLDEATSALDVESERIVQEALNRMMVERT
 TLVVAHRLSTVRNVDCITVVRKGKIVEQGPDALVKDPDGAYSQILRLQETHRDERHKL
 PDSRSKSTSLFRRSRTKDFLSKSNRYSFKSPLGLPVDIHEDGMTSEQQKVDHSDSKAIK
 KTFPGRFLNFKPEVPVLLLGSIAASVHGVILPLYGIIMPVGLKSFYEPPDQLRKDSRFA

Figure 7
 Page 4 of 32

LMSVVLGVACLISIPAEYFLFGIAGGKLIQRVRTLSFQRMHQEVAWFDKPSNSSGALGT
 RLSVDALNVRRLVGDNLALIVQAVATLITGFAIAFAADWRLALITCVIPLVGAQGYAQV
 KFLKGFSEESKEMYEDANQVAADAVGSIRTVASFCEKRVVAIYNKKCEALRKQGIRSG
 IVGGIGLSFSNLMLYLTYGLCFYVGAKFVSQGKTTFSDFVKVFFALVLAAVGVSQSSALS
 TNATXARDSAISIFSIDRKSRISSSDEGAIMENVTGSIIDFNNVSFKYPSRPDVQIFSDFTL
 HIPSQKTIALVGESGSGKSTIALLERFYDPDSGNISLDGVEIRSLKVSWLRLDQMGLVGQE
 PVLFNXTIRANITYGKHSEVTEEEITAVAKAANAHEFVSSLPQGYDTVVGEKGVQLSGG
 QKQRVAIARAILKDPKILLDEATSALDAESERVVQDALDRVMVNRTTIVVAHRLSTIKG
 ADMIAVLKEGKIAEKKGHEALLRIKDGAYASLVQLRSNSE*

>22825 OsPN22825

MAEAEAGRKEKKEVVRLERESVIPMKPKLIMKLAYLIEQQSDREEFLKLCKRIEYTIRA
 WYHLQFDDMMELFALFDPVHGAQKLQQQNFSTEEVDTLQNFILTYFFQVMQKSNFNIL
 SDDEVELAHSGQYLLNLPKIVDEAKLDNKLKSKYFKEHHHDNLPEFSKDYVIFRRGIGLD
 RTSNFFFMEKVDMIARAWRWFLEKTRLQKLFSRKKSVRPKTDSKKNDLV

>29041 OsPN29041

TNSLVQIRRALRALVDDHTDGLMDFEDTEVRSSSEETDALEEARLAIEQVVIPKGESVQLL
 PRPPSIASQVDLVESFKLKWESIGQEPNACLRLPQFVGVEEGGKSVKQEAATELTDSDN
 SDDMD

>29076 OsPN29076

ERDISMVNSYPPSAFDCKGWNSASIFIYESDEBIQCLLDWLRDYPREKELKDSILQWQR
 HFCHQSSSPLVDPPISGPKGEQLMELPNTKAAVILEQKYGLQLDQDTSPLPKRGKKIKL
 SSERTRYRCDCLEPVWPSRYHCLTCHETYLSTEFEGHNDGKCSKIHQSPDESRENDEPK
 VKVTKSDTKEKDSLECSSVIEPSSDRKLMQCPYDFEEICRKFTVNDNSKETVKQIGLNGS
 NGVPSFVPSPAFFLEPAIVQSQNXKDDELKDWTSLEECSNAMSQKLQEVSKSGQSCP
 GNVGDEKVQKSKKPTPDNTSGEEAHSTTGKPTRLLAVNGGLVPESLRPLGRNSHILKQ
 QKINLLDIEAALPEEALRASKCQQIRRRSW

>29077 OsPN29077

VEAQKKIEAAIROKIDENWEAALEHNPEAFARVVMPLYVDMEVNGVPLKAFVDSGAQ
 STIISKSCAERCGLRLLDQRYRGVAIGVGQSEILGRIHVAPIKIGHVFYPCSFVLDAPN
 MEFLFGLDMLRKHQCIIDLKDNVLRVGGGEVSVFPLQEKDIPSHIRDEEKLKSLASLSQG
 AAGESSTAREKTPDAPRAPTTGAPAVNPPQPQGGGDFEAKVTKLVELGFDRASVIQAL
 KLFNG

>29084 OsPN29084

ESGIFRQILRGKLDLES DPWPSISDSAKDLVRNMLIRDPTKRFTAHEVLCHPWTVDDAVA
 PDKPIDSAVLSRLKHFSAMNKLKKMALRVIAESLSEEEIGGLKELFKMIDTDNSGTITYD
 ELKNGLKRVGSDLMEPEIQALMDAADIDNSGTIDYGEFLAATLHMNKLEREENLVSAFT
 FFDKDGSGFITIDELSQACEQFGLSDVHLEDMIKDQNDGQIDYSEFAAMMRKGNA
 GGANAGGVSTSTGGTGR

>29086 OsAAB53810

MTLVKIGPWGGNGGSAQDISVPPKLLGVITYSSDAIRSI AFNYIGVDGQEYAIGPWGGG
 ESTSTEIKLGSSEQIKEISGTHGPVYDLADIVTYLKIVTSANNTYEAGVPNGKEFSIPLQDS
 GHVVGFFGRSGTLIDAIGIYVHP*

>29098 OsPIP2a

MGKDEVMESSGGAAGEFAAKDYTDPPPAPLIDAAELGWSLYRAVIAEFATLLFLYITV
 ATVIGYKHQTDASASGADAACGGVGLGISWAFGGMIFILVYCTAGISGGHINPAVTFG
 LFLARKVSLVRILYIVAQCLGAICGVGLVNAFPNAYFNRYGGGANTLAAGYSKGTGL

Figure 7
 Page 5 of 32

AAEIGTFVLVYTVFSATDPKRNARDSHVPVLAPLPIGFAVFMVHLATIPITGTGINPARSI
 GAAVIFNNEKAWHNHWIFWVGPFVGAALAAFYHQYILRAGAIAKALGSFRSNA*
 >29113 OsPN29113
 CKVSYIDAILGTTVKVPTVDGMVDLKIPSGTQPGTTLVMSKKGVP LLGKSNARGDQLVR
 VQVEIPKRLSSDERKLIEELANLNKAQTANSRR*
 >29115 OsPN29115
 VSFPYSPRPAALAAGARASRVSPVVVAAGGGHQRLMGS LTNTQGLRFGVVVARFNEIV
 TNLLLQGALET FERYSVKKENITVLSVPGSFEIPVAAQKL GKSGKFDANFVHWRC AFRG
 DTTHYGAVANSAASGVLSAGLSAEIPCIFGVLT CDDMDQALNRAGGKAGNKGAEAA LT
 AIEMASL FQHHLA*
 >29116 OsPN29116
 GRKPVA AQHYNKGDKDGTSTGSRCTSVAVVPEREGIFVVSHSDGNLYVYDKCKDGNT
 ECTFPAIKDPAQLMISHAKSSKSNPIARWHVCQGSIN AISFSPDGAYLATVGRDGYLRVF
 DFSKEQLIFGGKSYYGALLCCTWSSDGKYL LTGGEDDLVQVWSMDDRKTIVAW
 >29117 OsPN29117
 MLVQRRDGDGTGPAVRLRVSHGASFRDVAVPAHSTFGELKGVLTQATGVEPERQRLFFR
 GKEKSDNEFLHTAGVKDGAKLLLEKPAPANVEQRAEPVIMDESM MKACEAVGRVRA
 EVDRLSAKVCDLEKS VFAGRKIEDKDFVVLTELLMMELLKLDGIEAEGEARAQRKA EV
 RRVQGLVETLDK LKARNANPFSDQNKSVSVTTQWETFDNGMGSLNAPPPRV SSTQINT
 DWEQFD*
 >29118 OsPN29118
 TSSGDQQMVTV AERFPREV SSEAVFRCVRLGPVDQAEAEVAYQTA VSIGGHVFKGILHD
 VGPEALAVAGGGGASEYHFRLTGDGSSPSTAAAGEAGSGGGGN IIVSSAVVMDPYPTPG
 PYGAFPAGT PFFHGHPRP*
 >29119 OsPN29119
 MEFADGARWPEPRGDAAGAPPLERGDAPSPRFDSSRALRLLRELGS NVTEDLVVLMP
 NLLSFLKHDDPVVNQSIASGTNLFAAVLEEMTLQINKCGRVDAWLEEMWAWTKQFK
 DAVHNLIHESVPVATKLF AVKF IETWILCFAPQSKSDRMQPTEGRNRRLFDSSRLSQFHP
 SLNPAVLEADANRALILLVDILQSA CAHQGSFLVGTINSLAAIAKNRPVYYERILPVLLGF
 DPSLEVAKGAHPASLRYSLKTAFLGFLRSPCQAMIESKDTLVRQLRVLSPGEATEQIIRQ
 VEKMTRNIERASRASKDEPSTLDMPYGDVSR
 >19701 OsSSS
 MATAAGMGIGAACL VAPQVRPGRRLRLQVRRRRCVAELSRDGGSAQRPLAPAPLVKQP
 VLPTFLVPTSTPPAPTQSPAPAPTPPPLPDSGVGEIEPDLEGLTEDSIDKTIFVASEQESEM
 DVKEQAQAKVTRSVVFTGEASPYAKSGGLGDVCGSLPIALALRGHRVMVVMPRYMN
 GALNKNFANAFYTEKHIKIPCFGGEHEVTFHEYRDSVDWVFDHPSYHRPGNLYGDN
 FGAFGDNQFRYTLLCYAAACEAPLILELGGYIYGQKCMFVVNDWHASLVPVLLAAKYRP
 YGVYRDARSVLVIHNLAHQGV EPASTYPDLGLPPEWYGALEWVFPEWARRHALDKGE
 AVNFLKGAVVTADRI VTSQGYSWEVTTAEGGQGLNELLSSRKSVLNGIVNGIDINDW
 NPSTDKFLPYHYSVDDL SGKAKCKAELQKELGLPIRPDVPLIGFIGRLDYQKGIDLIK LAI
 PDLMRDNIQFVMLGSGDPGFEGWMRSTESGYRDKFRGWVGFSVPVSHRITAGCDILLM
 PSRFEP CGLNQLYAMQYGTVPV VHGTGGLRDTVENFNPF AEKGEQGTGWAFSPLTIEK
 NAVGIADGNFDIQGTQVLLGGSNEARHV KRLYMGPCRLTV*
 >20462 Os006819-2510
 MAFRLSNSLLGILNAVTFLLSV PVLGGGIWLATRADGTECERYFSAPVIAFGVFLLLVSL
 AGLVGACCRVNCLLWFYLVAMFVLIVVLCFTVFAFVVTNKGAGEAVSGRGYKEYRL

Figure 7
 Page 6 of 32

GDYSNWLQKRMENSKNWNRRSCLQDSKVCKKLQDKNWDRTQFFKADLSPLESGCCK
 PPSSCNFLYVSGTNWTKVPTNSSDPDCNTWVDDGTQLCYNCQSCKAGAVATLKRDKW
 RVAVVCIVFLVFIVIVYSLGCCAFRNNRRDNRGAYRGAAWKGGYA*
 >20544 OsCRTC
 MAIRARSSSYAAAAVALALALASVA VAGEVFFQEKFEDGWESRWVKSEWKKDENM
 AGEWNHTSGKWNNGDPEDKGIQTSEDYRFYAISAEYPEFSNKDKTLVLQFSVKHEQKLD
 CGGGYVKLLGGDQKKGFGDTPYSIMFGPDICGYSTKKVHTIFTKNDKNHLIKKDVPC
 ETDQLSHVYTLIHPDATYSILDNVEKQSGSIYEHWDILPPKQIKDPEAKKPEDWDDKEY
 IPDPEDKKPEGYDDIPKEIPDPAKKPEAGADEEDGEWTAAPTIRNPEYKGPWKQKKIKNP
 NYQGWKPPMIATPDFQDDPYTYAFDSLKSIGIELWQVKSGLFDNITDDAALAKTFA
 EETWAKHKDAEKAADFDEAEKKKEEEEEAAKAGEDDDDLDDDEDAEDEDKADEKADSDA
 EDGKDSDEKHDEL*
 >20551 Os003118-3674
 MGGKELSEEQVASMREAFSLFDTGDGRIAPSELGVLMRSLGGNPTQAQLRDIAAQEKL
 TAPDFDPRFLDLMLRAHLRPEPFDRPLRDAFRVLDKASGTVSVADLRHVLT SIGEKLEPH
 EFDEWIREVDVAPDGTIRYDDFIRIVAK*
 >20554 OsRAB16B
 MENYQQGHGYGADRVDVYGNPVGAGQYGGGATAPGGGHGAMGMGGHAGAGAGGQ
 FQPAREDRTKGILHRSGSSSSSSSEDDGMGGRRKKGIKEKIKEKLPGGNKGNNQQQQ
 QMMGNTGGAYGQQGHAGMTGAGTG VHGAEYGNAGEKKGFMDKIKEKLPQQH*
 >22883 OsLIP5
 MAGIIHKIEKLHMGGGGEHKKKEDEHKKKEGEHHKKDGEHKEGVVEKIKDKITGDHGDGG
 EHKEKKDKKKKKKEKKHGEEGYHHDGHSSSSSDSD*
 >23226 OsPN23226
 IKRLHLLTVKESAMDVPTNLDARRRISFFANSLFMDMPSAPKVRHMLPFSVLTPYYKE
 DVLFSQALEDQNEGDVSLFYQLKIYPDEWKHFLQRVDCNTEELRETEQLEDELRLW
 ASYRGQTLTRTVRGMMYYRQALVLQAFLDMARDEDLREGFRAADLLNDESPLLTQCK
 AIADMKFTYVVSQCQYGIQKRSQDHRAQDILRLMTTYP SLRVA YIDEVEEPSKDRNKKI
 EKVYYALVKA AVTKPDDPGQKLDQDIYRIKLPGNAMLGEGKPENQNHAIIFTRGEGLO
 TIDMNQEHYMEETLKMRNLLQEFLKKHDGVRYP SILGVREHIFTGSVSSLAWFMSNQ
 ETSFVTIGQRVLANPLRVRFHYGHPDIFDRLFHLTRGGVSKASKIINLSEDFAGFNSTLRE
 GNVTHHEYMQVGKGRDVGLNQISLFEAKIANGNGEQTL SRDVYRLGHRFDFRMLSCY
 YTTIGFYFSTMMTVWTVYVFLYGRLYLVLSGLDEALATGKRFIHNEPLQVALASQS FVQ
 LGFLMALPMMMEIGLERGFRTALSDFVLMQLQLASVFFTFSLGTKTHYYGTLLHGG A
 EYRATGRGFVVFHAKFAENYRLYSRSHFVKGIELLILLIVYEIFGQSYRGAIA YIFITFSM
 WFMVVTWLFAPFLFNPSGFQWQKIVDDWTDWNKWISNRGGIGVPPEKSWESWWEKEQ
 EPIKYSGRKRGIVLEIVLALRFFTYQYGLVYHLNITKHTKSVLVYCLSWVVFVILLVMKT V
 SVGRRKFSADFQLVFRLIKGLIFITFISIIILIAIPHMTVQDIFVCILAFMPTGWGLLLVAQAI
 KPVIVRIGLWGSIKALARGYEIIMGLLLFTPIAFLAWFPFVSEFQTRMLFNQAFSRGLQISR
 ILGGHKKDRATRNE*
 >23485 OsPN23485
 HASGSGSVRERFEAMIRRVQGEVCAALEEADGSGARFVEDVWSRPGGGGGISRVLQDG
 RVFEKAGVNVSVVYGVMPD AYRAAKGEAGKNGAAADGPKAGVPFFAAGISSVLHP
 KNPFAPTLHFNYRYFETDAPKDAPGAPRQWWFGGGTDLTPSYIEEDVKHFHSVQKQAC
 DKFDPSFYPRFKKWCDYFYIKHRNERRGLGGIFFDDLNDYDQEMLLNFATECADSVV

Figure 7
 Page 7 of 32

PAYIPIIERRKDTPTFTEEHKAWQQLRRGRYVEFNLVYDRGTTFGLKTGGRIESILVSLPLT
ARWQYYHTPEEGTEERKLLDACINPKEWLDL*

>23878 OsPN23878

EXGGGXAEVDSAAAATGARATAATTPRRQSPAGAGSPSQRDARWGDTSSYGARKKHR
VFCQLPNSDWALCTVITTSGDSDVLKLPPEGKVLRLKTESLEAANPEILDGVDDLMQLSY
LSEPSVLYNLQYRYSQDLIYTKAGPVLVAVNPFKKVPLYGNEYIDAYRNKTKDSPHVYA
IADSALREMKRDEVNQSIISGESGAGKTETAKIAMQYLASLGGGGGIEYEILQTNPILEA
FGNAKTLRNDNSSRFGKLIHFSTTGRICGAMIQTFLLEKSRVVQCAVGERSYHIFYQLC
AGAPASLRDKLNMKKADEYKYLKQSCCYSIAGVDDAQMFRTVTEAMNIVHISKEDQD
NVFTMVSAILWLGDVSTVIDNENHVEIVVDEAAETVARLLGCSIEDLNLALSKRHMKV
NNENIVQKLTLSQAIDTRDALAKSLYASLFEWLVEQINKSLVVGKRRTRGRSISILDIYGFE
SFDRNSFEQFCINYANERLQQHFNRHLFKLEQEEYVEDGIDWAKVEFEDNQNCNLFEK
KPLGLLSLLDEESTFPNATDLTFANKLKQHLNNNSCFRGERGKAFVRHYAGEVAYDTS
GFLEKNRDLLHMDSIQFLAKCKSSLPQMFASKMLSQSDNPLPVPYRNSAADSQKLSVA
MKFKGQLFQLMQRLESTTPHFIRCKPNNLQLPAIYEQGLVLQQLKCCGVLEVVRISRG
YPTRMTHQKFARRYGFLLEDVASQDPLSVSVAILHQFNILPEMYQVGYTKLFFRTGQI
GKLEDTRNRTLHGILRVQSCFRGHQARRHARERIRGV LALQSFIRGENARKMYSSLARK
HRAAILQRNLKCWLARRYFVNIRKASVVIQSGIRGCLVRRCAGNVDLLNVLREFESKK
EAEGDQILIKASFLAELQRRILKAEATVREKDEENEMLQORLQQYENRWSEYEQKMKA
MEEMWQKQMRSLQSSLSVAKKSLALDETPRMSDSSVDQSWESNGVHIGSASQLVPRTV
GREMNASISVISRLAEFEQRSQVFADDAKFLVEVKSGQADASLNPDMELRRLKQNFDS
WKKDFGSRIRETKVILNKLGSNESSPNSVKRKWWGRLNTSKFS*

>24059 OsAAK01712

MDASAGGGGNSLPTAGADGAKRRVCHFYDAEVGNYYCGQGHPMKPHRIRMTHALLA
HYGLLDQMQLKPHPARDRDLCRFHADDYVAFRLSVTPETQQDQIRALKRFNVGEDCP
VFDGLYSFCQTYAGGSVGGAVKLNHGHDIANWAGGLHHAKKCEASGFCYVNDIVLAI
LELLKYHQRVLYVDIDIHGHDGVEEAFYTTDRVMTVSFHKFGDYFPGTGDIRDIGYSEG
KYYCLNVPLDDGIDDDSYQSIFKPIISKVMEMYPGA VVLQCGADSLSGDRLGCFNL SG
KGHAECVKFMRSFNVPLLLLGGGGYAIRNVARCWCYETGV ALGHELTDKMPPN EYFE
YFGPEYSLFVAASN MENRNTNKQLEEIKCNILDNL SKLQHAPSVQFEERIPETKLPEPDE
DQEDPDERHDPDSMDVLDHDKPMGHSARSLRNIEVKREITESEAKDQH GKRLTTEHKG
PEPMADDPGSSKQAPVSRRLLYPSANP*

>29037 OsPN29037

MEVGFLGLGIMGKAMAANLLRHGFRVTVWNRTLSKCQELVALGAAVGETPAAVVAK
CRYTIAMLSDPSAALS VVFDKDG VLEQIGEGKGYVDMSTVDAATSCKISEAIKQKGGAF
VEAPVSGSKKPAEDGQLVILAAGDKVLYDDMVPAFDVLGKKSFLLGEIGNGAKMKLVV
NMIMGSMMNALSEGLSLADNSGLSPQTLLDVLDLGAIANPMFKLKGPSMLQGSYNPAF
PLKHQQKDMRLALALGDENAVSMPVAAASNEAFKKARSLGLGLDLDFAVYEV LKGAG
GSGKA*

>29950 OsPN29950

ASEESTTANLVGQITTTCTDDISVNRSAENSSQKNIPLDGVSAQSIKPSASSRSEPVGLGG
GLQPKRRSR TAKPPGSSSDTGEVVNSSRISNSQNAVSMGQQVLQALASQNTNVNRSHVT
DSPLPSTTSQFSGGMPPRRQGGEGQVDFGSMISSVLNNPAFGNLLSNVAEQTGMGSAGD
LRNMVEECAQSPAIMDTMSNLVQNV DGSGRGQGGIDLSRMMQQMMPVVSQVLGGAG
ARPAGTNSGQSRLQPRRSDMRVDDASDYGNSQIDLHQAREHIEQHDSPRDIFGAVLETA
AQAYGEDESIEDMLEELVSDPELTDDY LKLLLQQVRQRIQSASQSGNQ S*

>20621 OsAAK38313

MHMMRRLKSIASGRSSVSDPGGDSGTRPKFDQDGAGDIVIEPHLTDDKPMRVDQESS
SHRDAEASTSTSKNPGRTTEEAGADILPKEMNEMTISDDKVDGHNDKESEGVIVNANGTE
TGQIVTSIGGQNGKPKQKVS YMAERVVGTGSFGVVFQAKCLETGETVAIKKVLQDKRY
KNRELQTMQLLDHPNVVQLKHFFSTTERGEVYLNVLVEYVSETVYRVAKYYNRMNQ
RVPILHVKLYAYQMCRALAYIHRVVGVCCHRDIKPQNLLVNPHTHQLKLCDFGS AKKL
PGEPNISYICSRYYRAPELIFGATEYTTAIDIWSVGCVLAE LLIGQPLFPGESGVDQLVEIK
ILGTPTREEIRCMNPNYSEFKFPQIKAHPPWHKLFGKRMPEAVDLVSRLLQYSPNLRCTA
VDACAHPPFDEL RDPKTCLSNGRSLPPLFDFSAAELEGLPVELVHRIIPEHMRK*

>25962 OsAAK18843

MVFKGRFFSSRHKSSESSSPTDGSNSPRTPTSAPPAAGSPAAASSSSSSSRSDKKKPKSETP
RKRDKLFGGSASGSAASKGGAGAAPASASSSPSADGRKAAAAQLRDGGAGGASAAAL
SPILASSLGLNRIKTRSGPLPQEGHRIAAALGSSNLSRGQAQADPSAASAGGGGGGGRKA
GSSWADSTSGSRGKGKAAEHPARGATATSLEGKSSAKDVLSYKEIGAHDFFHFAAWCLD
AKLGTGANCAYPDCEPKESESPRFKAIMQATSAPRKRVPADIKSFSHELNSKGVRPFP
FWKPRGIYNLKEVLKVIQVRFEKAKEEVNSDLAVFAGDLVGVMKEYADSHPEWKETLE
DLLILARSCCVMTPEGFWLQCEGIVQDLDHRQELPMGV LKKLYTRMLFILTRCTRLQ
FHKESGFAEDEVMQDRDKIISADRQLAQPGDDTTTRGSKSDVRKSYSQEQHNLKW
KRSQEI KPVKFLSPLD TTDVKKEVESPTREIRISSWKPFPSVPKPKDP TPIKEESPNKKT
TPPAVSSQAE LNSPVESTSHQSLPPKHQHKTSWGHWS DQPNISEEGSIMCRICEEYVPTH
YVENHSAICASADRC DQKGVSVDERLIRVAEAEKL VESYTQKDL PNAV GSPDVAKVS
NSSINEESDGPSPKLS DWSRRGSADMLDYLQEADSTISLDDIKNLPSMTCKTRFGPKSDH
GMATSSAGSMTPRSPLTTPRSNHIDMLLAGRSAINESDDL PQIVELADIARCIATPLDEE
RALSLVTCIEDLQEI VNRKHEALTVQTFGTRIEKLHREKYLLLCDSVDMDKVDSASTV
MDEEDDVVRSLRASPVHPVKDRTSIDDFEIKPISRGAFGRVFLAKKRTTGDLFAIKVLRK
ADMIRKNAVESILAERDILITVRNPFVVRFFYSFTSRENLYLVMEYLN GGDLYSLLRNLG
CLDEDVARIYLA EVVLAEYLHSMHIVHRDLKPDNLLIAHDGHIKLTDFGLSKVGLINST
DDLSGPAVSGSSLYGDDEPQMSEFEEMDHRARRQKRS AVGTPDYLAPELLGTGHGTS
DWWSVGVL FELIVGIPPFNAEHPQTIFDNILNRKIPWPHVPEEMSSEAQDLIDKLLTEDP
HQRLGANGASEVKQH QFFKDISWDTLARQKA AFVPSSDSAFDTSYFTSRYSWNPSDENI
YEA YEFEDSSDNGSLSGSSSCVSNHQDDMGDESSGFTEFESSNVNYSFSNFSFKNLSQL
VSINYDLLTKGLKDDPPTKSET*

>27024 OsBAA85416

MYSSPKALLYLFLH LAYLYRRLGPTSRTPHFGPGSNHPAHFMFTPSLGRLRALRVPCGPSA
HVAERQTPKQTD TNLNAAAPPASRPASDAGAAAAAGDPPLAAAGCHDRPDGRRAVEV
PPRHRPAGIPQTPTAKPFLPSSTASRPPPPPLSQGVGKTTLVTRVFETLRESHPHLNIR
GFYTREVRESGERVGF EVVTLDGRTGPLASSKVSSRESVRWPTVGRYKVDIASLES LALP
ELQVKDDTDLFIIDEV GKMELFSSAFFPAVMRVIESNIPVLATIPVPRLGRDIPGVARLR
NHPGAVIYTLNTGNRDAMREGVYNHLSLLQKRNAV RIMSRRRSNMCKHVSFKLNKL
CFSVLFILQHQAPTSGSDLHLHKIQSTWMKHRRQGGENMGPSLLIACFSWQFGGIIAKH
FMAFSSSPACRQQLARMP SQEEIKASLPASVKNLQERDP ISTVGEATVSPLIGGGRPVV
AGELHANCREEAREDP SYCWISLATSATMLQYLD FSHASTSRKWSHKKQEGEFAPRN
SMEFTLEAPQSYGVFQEDVPYSCNMRQQYPKAGLNHSSSPIKRIIHEDVSFR TNEVQKRP
SVIARLMGMDSPPMSTTAGELAAAGHTEEK RQDMITSTRPMPRRDPSEMVSTKHVSFVQ
HKGSMKHSPKQAEVCA YDDSMELFGQLSKAISSEWAKPQPREHPQEEELQFKKDFE
AWQASRMWEQSR ALELESHLDDDDVRCTDIVPYRFQHRGKDNAGKKHTHSNGDAH W

Figure 7

Page 9 of 32

RRSKESGTGTGTSISGSRTFSLTSADASSTRPLSRFYEEERLLSPKKIVILKPCPEMSTD
 DIESSLGSPPEMVKKENNMEAFLEEVKKRLKVELEGRMASDDRAADRWAAGGDIPADP
 KQIARSIANQIRETVTKDLHSALLRSESTRSYRSDIPLNGQSQMDYICRDARKHLSRLK
 NVLRREPETEPALSHRRRTASASFNEEPRPKPRHEVARKGKIRSKEEKHAIEFDVRSFR
 RGHKKASPTPAIDSDPVSPRNLRSFSAPVSGTTFVKLLSEEPRVLTGARLQKQEGYGSR
 PPPPSSEEERKGRKDTFNIKGRVSNLRQNLGLRAKLFGKKLHSADESPFPDDLPPIGTLVT
 APSVLIHPGVLQENSTEVPSPASWCSSPPDEMSSRGGYPSVSPLEASFSEHRSPLKMAAR
 DMSSSASEPEHPSSEQAQTDQELAETSPIQDDDDDDTDEIDNPIKAYIRAILVIAGLYGQR
 RSSDQLFSDREVKPIPAWVFEEVESSSSSSAPATTDCDAAATGVDRHLLFDLINESLPRVV
 QSSTTLCAFSRWYGAAPRRSPGGKRLLDGLWNTVQAWLAPPPPTDSPNSVDELIGRDM
 SMSPWNGPFREDVGAAGAEMEAELDELVDETLWDVLLNVGD*
 >23221 Os003316a-P014
 MARRSPCLAVAMLLLGALAVASAFIDEAAAAGRGLGHGARFMSKQGRVITYEKLPEPEP
 KPKPKPHKPTPKPEPKPEPEPKPVPEPEPKPEPKPEPKPEPKPYPEPKPEPKPEPKPE
 PEKPEPKPEPKPEPKPYPEPKPEPKPEPKPEPKPEPKPKPEPKPHPEPKPDPEPKPHPEP
 EPKPEPKPEPKPHPEPEPKPEPKPEPKPEPKPEPKPEPKPKPKPEPKPKPEPKPYPEPKPKPE
 PKPEPKPEPKPEPKPEPKPEPKPEPKPEPKPKPEPKPHPEPKPEPKPEPKPEPKPEPKPEP
 KPEPEPKPEPKPEPKPEPKPYPEPKPDPEPKPHPEPKPEPKPKQPEPKPEPKPEPKPEPKPE
 PKPEPKPYPEPKPEPKPKPKPEPKPEAPPKHKPHIPPATDQ*
 >20115 Os011374-4328
 MVCVAIEYRMRRGQRDRAPASADEERGS DSSSSSSDDVTEDDRRGSCTDCVSIKHL
 ESANTMFSFIWWIIGFYWISAGGEDVIRDAPQLYWL CIVFLAFDVFFV VFCVALACIIGIA
 VCCCLPCIIAILYAVSDQEGASEDDIRQIPRYKFRRTDEPEKQTADETGPFGGIMTECGTN
 QPIEKVLAPEDAECCLISAYDDGAELRELPCGHFHFCACIDKWLHINATCPLCKFNIRK
 SGSSSGSEEV*
 >20285 OsSGT1
 MATAAASDLESKAKAAAFVDDDFELAAELYTQAIAS PATAELYADRAQAHIKLGNYTE
 AVADANKAIELDPSMHKAYLRKGAACIRLEEYQTAKAALELGYSFASGDSRFTRLMKE
 CDERIAEELTEVPVKKAEDGAAAPSVASFVEEKDDAANMDNTPPMVEVKPKYRHDFYN
 SATEVVLTIKAGVPAENVVVDVFGEQMLSVSIEVPGEOPYHFQPRLF SKIPEKSRYQVLS
 TKVEIRLAKAEQITWTS LDYDKKPKAVPQKIIPPAESAQRPSYPSSKSKKDWDKLEAEVK
 KEEKEEKLEGDAALNKFFRDIYSDADEDMRRAMMKS FVESNGTVLSTNWKDVGSKKV
 EGSPPDGMELKKWEY*
 >24060 OsPN24060
 MASA AVIPSSAPGAAAAGAAAVXSAGWVVDERDGFISWLRGKFAAANAIDLLLLXLR
 SVGEPGEFEHVAAA VQQRHHPVAPVIHMQQFFPVGDVAYALQQAGXRRRAPPHHQQQ
 GPGASPSPPPXPGRPSFSASHSHRHGGHHRS DSVRGGGTGATAGSDKDGREHVHNK
 EEKGMKEAENVVEAKSSQLESLSHEGEKTPRPQAVAEGSSKVVPPTPVEYTVNDIIDGK
 TVNAVEGLKVYEGLVNENEKNKILSLLNETKASFRRGGLEAGQTVIIGKRPMKGHGREGII
 QLGPIVEGPPEDDYPRETKVEAVPGLLHDLFDRLCQKEIPTKPDYCVIDYYNEG DYSHP
 HQSPPWYGRPFCTFCLTDCDMVFGRVISGERGDHRGPLKLLLSTGSLLV LHGKSADVAK
 RAIPAACKQRILLSFGKSLSRKQVPSES VSRFTTPLTPPPMPWGPPR PANMARHSSSPKHF
 GYAPNSGVLPAPAIGAHHIPPSDGMQPLFVAPAPVAAAAMPFPSPVPLPNLTNSLDGRSC
 SKVCSTTTTPSSWHGGLPSSWIGSRTASSDD
 >23914 OsPN23914

Figure 7
 Page 10 of 32

MEGGGEVGVYVLGPNQEHVGPYALSELREHFANGYISESSMLWAEGRSEWMPLSSIPD
LLAVVTKKDQPDEGIEDDFDKFQKEVIEAEAEVEASTDKAADNDINQEHGADDPDDR
ATPPDGEDEFTDDDGTVYKWDRVLRAWVPQDDLEGKNDNYEVEDMTFAHEEEVFQAP
DIAGSTTLEENNVSAEIEIKEPPKVEKRAHKKRKSSEKPADKKEAYKPPDSWVDLKVNT
HVYVTGLPDDVTAEEIVEVFSKCGIikedPETRKPVRVKIYTDRETGRKKGDALV
>24061 OsPN24061

MAGADVDVGTRELRLGLPGGGGGAAEAAAKAARGFEETIDLKLLPTAGMEEAAAGK
AEAPAAEKAKRPAAEAAAADAEEKPPAPKAQAVGWPPVRSFRRNIMTVQSVKSKKEEAD
KQQQQPAANASGSNSAFVKVSMGAPYLRKVDLKMYSYKDLSSLALQKMFGTFTAT
GNNMNEVNGSDAVTTYEDKGDWMLVGDVPWQMFVESCKRLRIMKGSEAIGLAPRA
KDKYKNKS*

>24063 OsAAB28535

MNPEYDYLKLLIGDSGVGKSCLLLRFADDSYLESYISTIGVDFKIRTVEQDGKTIKLOI
WDTAGQERFRTITSSYYRGAHGIIIVYDVTDQESFNNVKQWLNEIDRYASENVNKL
GNKCDLAENRVVSYEAGKALADEIGIPFLETSKDATNVEKAFMTMAGEIKNRMASQG
RTNASKPATVQMPRPVAAQQSSCCS*

>28982 OsARCN1

MVVLAASIISKSGKALVSRQFVDMRIRIDGLLAAPFKLVGSGKQHTYVETENVRYVYQ
PIEALYLLITNKQSNILEDLDLRLLSKLIPEYAPSLDEEGVCKAAFELLFAFIEAISLGK
ENVTVAQVKQYCEMESHEEKHLKLVMSKINETKDVMMRKVTEIEKSKTDRGKPKDG
GFGSLRTPNSFSMDGIRGGGPGDPIFGDMDSFTHKAKGRPSAPAPASASTKVPGGMKL
SKAQKTNQFLESKAEGEVILEDTQPSATQSRSSYIPSPDITVTIEEKLNVTVKRDGGVS
NFDIQGTLALQILNDTDGFIQLQIENQDVPGLNFKTHPNINKELFNSQQIVGAKDPNRPFP
SGQNETPLVKWRIQELNESSLPLAVNCWPSVSGNETYVNIEYEASEMFDLHNVVISIPLP
ALREAPGVRQIDGEWRYDSRNSVLEWSIILVDQSNRSGSLEFTVPAADPSTFFPISVGFA
SNTFSDLKVTAIRPLREGSPPKFSQRNRLVTYNYQVV*

>29042 OsPN29042

MAGSSSLSTLSLCTQSPSPSPVASGRLVAPAVLGFAAPRFTLRAAPRRLTARAVAGDA
EDEXGKEPAADQGGAAA VAEAPADVPTSEVAELKAKLKEALYGTERGLRASSETRA
EVVELITQLEARNPTPAPEALTLLNGKWILAYTSFSQLFPLLGGGSLPQLVKVEEISQTID
SENFTVQNCIKFSGPLATTSVSTNAKFEVRSPKRVQIKFDEGIIGTPQLTDSIVLPEKFELF
GQNIDLTPLKGIFSSIENAASSVARTISGQPPLKIPRTDNAESWLLTTYLDDELRI SRGDGS
SIFVLKEGSTLLY*

>23949 OsPN23949

TRPDFDKKVVDWLASNFKKDEGIDLLKDKQALQRLTEAAEKAKMELSTLSQTNISLPFI
TATADGPKHIETTL SRAKFEELCSDLIDRLKTPVTNALRDAKLSVDNLDEVILVGGSTRIP
SVQELVKKITGKDPNVTVPDEVVSLGAAVQGGVLAGDVKDVLDDVTPLSLGLETLG
GVMTKIIPRNTTLPTSSEVFSTAADGQTSVEINVLOGEREFVRDNKSLGSFRLDGIPSLH
HVVFHKLKSSLILMPMVSLLLLLLIRVLGRNRISPSPALSTLPKDEVERMVEEADKFAQE
DKEKRDAIDTKNQADS VVYQTEKQLKELGDKVPAPVKEKVD AKL NELKEAIAGGSTQS
MKDAMAALNEEVMQIGQAMYNQQPNAGAAGPTPGADAGPTSSGGKGPNDGDVIDAD
FTDSN*

>20696 OsERG3

MVQGTLEVLLVGAKGLENTDYLCNMDPYAVLKCRSSEQKSSVASGKGSDPEWNETFM
FSVTHNATELIKLMDSDSGTDDDFVGEATISLEAIYTEGSIPPTVYNVVKEEEYRGEIKV
GLTFTPEDDRDRGLSEEDIGGWKQSS*

Figure 7
Page 11 of 32

>29983 OsPN29983

GVETVVTDTAGNKVVVTGAADAAELKERIEARTKKA VQIVSAGAGPPPKKDKKEKKDK
DKKGGGDDKKA EKEKGGGGGDKKAEKEKGGGDKPKKEKKAKEPKEETVTLKIRLHCE
GCIDRIKRRICKIKGG

>30844 OsPN30844

VAVRKKETRDGAMQTMPSRVQQGAADFEETLFVRCHLYCSGGAGTGKPLRFEP RPFL
SAVAVEAPELDFGRSAVDLSLLVKESTDKSQQGERVRQWDMALPLAGKAKGGELVVK
LSFQIMDDGGVGLFNQTGAATKINSSSSSSSLFARKQSKLSFSITSPKVSRSSEPCLTPTKGS
PSPDLRGIDDFKLDEPSLPSLAEAKQEKEPEPEPEEKVDDSEFPEFDVVDK

>30868 OsPN30868

LDEVMAVSPVGLGRRSRQIFDEVWRKFSRLGQMSSASSTALAEQQAVLIRGGPMCEFT
VPGAQDTTVLVVGATSRIGRIVVRKLMLRGYNVKALVRRNDAEVIDMLPRSVDIVVGD
VGDPSTVKS AVSGCSKIYCATARSTITGDLNRVDNQGVRNVSKAFQDYYNELAQLRAG
KSSKSKLLIAKFKSPKSLNGWEVDQGSYFPNTFASRFDEGIDASFDSEAGQAVFSGFVF
TRGGYVEISKRLSLPLGSTLDRYDGLLFSVGGNGRSYVVILETGPLADTSQSKKYFARM
TKVGFCRVRVPFSAFRPVNPQDPPLDPFLVHTLTIRFEPKRQRPGDGSQSATDPRNFELIL
EYIKALPTGQETDFILVSCSGSGIEPNRREQVLKAKKAGEDALRRSGLGYTTVRPGPLQEE
PGGQRALIFDQGNRISQGISCADVADICVKALHDSTARNKSFVVCYEYVAKQGNELYEL
VAHLDPKANNYLTPALSVLEKNT*

>24292 OsBAA78745

MIRAIRACKTAAEERAVVRRECAAIRAAISEGDQDYRHRNMAKLMFIHMLGYPTHFGQ
MECLKLIAAGFPEKRIGYLGLMLLLDERQENAMFCFEHLVVLNADRDLNHSNQFIVGL
ALCALGNICSAEMARDLSPEVERLLQSREPNTKKKAALCSIRIVRKVPDLAENFMGSAVS
LLKEKHHGVLISAVQLCAELCKASKEALEYLRKNCLDGLVRILRDVSNSSYAPEYDIAGI
TDPFLHIRVLKLMRILGQGDADCSEFVNDILAQLCSVSYIELFALHFQVATKTESNKNAG
NAILYECVETIMGIEATSGLRVLAINILGRFLSNRDNNIRYVALNMLMKAMEVDTQAVQ
RHRATILECVKDADVSIRKRALELVYLLVNDANAKSLTKELVDYLEVSDQDFKDDLTA
KICSIVEKFSQDKLWYLDQMFKVLSLAGNYVKDDVWHALIVLISNASELQGYSVRSLYK
ALLACGEQESLVRVAVWCIGEYGEMLVNNVGMLDIEEPITVTESDAVDAVEVSLKRY
ADVTTTRAMCLVSLKLSSRFPTSESSLERMPVIDEASYLAKRAASTQATISSDKLAAA
ATPGSSLKL PNGVAKPPPAPLADLLDLSSDDAPATTSAPTAPNDFLQDLLGIGLTDST
AGFDLTFLFIQYIFGLTSEASDAHILTFDTHLCGAPSASTDILMDLLSIGSSPVQNGPPTVS
NFSLPQDENTAYPPITAFQSAALKITFNFKKQSGKPQETTHASFTNLTSNTFTDFIFQAA
VPKFIQLRLDPASSNTLPASGNDSVTQSLSVTNNQHGGKPLAMRIRITYKVNDKLIMNRR
RCSTKYDLVKHSSLLAIEAINPPLYRSERALVHEYLACVYRFRPFRNAPVQAIRRDVAEI*

>30845 OsPN30845

GDQGRTYLFRVSNVGVKTSVNVRIQGHSLRLVEVEGTHPVQNVYDSL DVHVGQSVAF
VTLDKA

>29997 OsPN29997

MGSLTRAEEEEETAAAEWSGEAVVYVNGVRRVLPDGLAHLTLLQYLRDIGLPGTKLGC
GEGGCGACTVMVSCYDQTTKKTQHFAINACLAPLYSVEGMHIITVEGIGNRQRGLHPIQ
ERLAMAHSQC GFCTPGFVMSMYALLRSSEQPPTEEQIEDSLAGNLCRCTGYRPIIDAFR
VFSKRDDL YNNSSLKNADGRPICPSTGKPCSCGDQKDINGSESSLTPTKSYSPCSYNEI
DGNAYSEKELIFPELQLRKVTSCLKNGFNIRWYRPLKLKQVLHLKACYPNAKLIIGNS
EVGVETKFKNAQYKVLISVTHVPELHTLVKEDGIHIGSSVRLAQLQNFLRKVILERDSH
EISSCEAILRQLKWFAGTQIRNVASVGGNICTASPIDLNPLWMATGATFEIDVNNNIRTI

Figure 7
Page 13 of 32

PAKDFFLGYRKVDLKPDEILLSVILPWTRPFEFVKEFKQAHRRREDDIALVNAGMRVYIRK
 VEGDWIISDVSIYGGVAAVSHRASKTETFLTGGKWDYGLLDKTFDLLKEDVVLAENAP
 GGMVEFRSSLTLSFFFKFLLHVTHEMNKGFWDGLHATNLSAIQSFTRPVGVGTQCYE
 LVRQGTAVGQPVVHTSAMLQVTGEAEYDDTPTPPNTLHAALVLSTKAHARILSIDASL
 AKSSPGFAGLFLSKDVPGANHTGPVIHDEEVFASDVVTCVGGQIVGLVVADTRDNAKAA
 ANKVNIEYSELPAILSIEEAVKAGSFHPNSKRCLVKGNVEQCFLSGACDRMEGKVQVGG
 QEHFYMEPQSTLVWPVDSGNEIHMISSQAPQKHQKYVANVLGLPQSRVVCKTKRIGG
 GFGGKETRSAIFAAAASVAAAYCLRQPVKLVLDRLDMMMTTGQRHSFLGKYKVGFTDDG
 KILALDLVDYNNGGHSHDLSLPVLERAMFHSNDVYDIPNVRVNGQVCFTNFPSTAFRG
 FGGPQAMLIAENWQHMA TELKRSPEEIKELNFQSEGSVLHYGQLLQNCIHSVWDELK
 VSCNFM EARKAVIDFNNNNRWRKRGIAMVPTKFGISFTTKFMNQVLRMNNCRLCLVQ
 VYTDGTVLVTHGGVEMGQGLHTKVAQVAASSFNPLSSIFSETSTDKVPNATPTAASAS
 SDLYGAAVLDACQQIMARMEPVASRGNHKSFAELVLACYLERIDLSAHGFYTPDVGFD
 WVSGKGTFFYYFTYGAFAEVEIDTLTGDFHTRTVDIVMDLGCSINPAIDIGQIEGGFIQ
 LGWAALEELKWGDDNHKWIRPGHLFTCGPGSYKIPSVNDIPLNFKVSLKGVLPKVIH
 SSKAVGEPPFFLGS AVLFAIKDAISAARAEEGHFDWFPLDSPATPERIRMACVDSITKKFA
 SVYYRPKLSV*

>30843 OsPN30843

GAGLQNLGNTCYLNSVLQCLTYTEPFAAYLQSGKHKSSCRTAGFCALCALQNHVKTAL
 QSTGKIVTPSQIVKNLRCISRSFRNSRQEDAHELMVNLLES MHKCCLPSGVPSSEPSAYE
 KSLVHKIFGGRLRSQVKCTQCSHCSNKFDPFLDLSLDIGKATSLVRALQNFTAELLDDGG
 EKQYQCQRCKKVVAKKKFTIDKAPYVLTHLKRFSFPNPREKIDKKVDFQPM LDKPFI
 SDSKVSNLKYSLYGVLVHAGWNTQSGHYCFVRTSSGMWHNLDDNQVRQVREADVL
 RQKAYMLFYVRDRVGNPTPRKDNITANMPARRTIPEKISGLSDMIQSGVIEAKLNGSSSP
 YGDKRLHGISNGNSIKTSREHYLKKDGKTEAPKASENNGLASTQKASAPQIDGATLSAQ
 SKQITSTGHREVSSSDRSASLTHVIVNQAVAMVPSQELQPKVDGLTDTSSLGNGNAILSE
 RNKQTSQHQNPFMSMPASHGKDTGAGLAAQTFTPKDAIVSNGVVPSSRDPISSEKVCGLQ
 KSIKQDDKT VKELPISENNIVSGLERNARKQTSSEVSMKVAAADSCNSNTPKRVDLKS
 KKLVRYPVMNMWLGPQVMLGSLKVQKKKKCNRTRRRSVVCEDMANATCSGNNTSE
 QQASTSTTTSSËTVQCTPRGRKRAYDSDSPKNNNQKQNKQDVIGADTGSGELNMDKRN
 VISETAASAELPKLGPSSANQEHSRNNVHAKLGVP RHFTVLTRDLAEVTVPCWDDVA
 VSNAEARESKHSEKSIGYVLDEWDEEYDRGKTKKIRNSKEDYGGPNPFQEEANYISQR
 NMKQRTYQPKSWKKHAHVRR*

>30857 OsPN30857

QDTRPLQAQRRRHGGGPQDRQREPSQGDFLRRQREEHPGGEEDRAPHGAGGHAAAGE
 GRGPRAHEHPQHQQGAAGAVGGGREGGGRAHLLRPRHRDIGDRVGHTYCLSRLRRIA
 QNEQPSIHPCPSLAGQVIIRWEFRIGRDNNTFQEKRRVFRGSSGFRM*

>20257 OsCYCOS2

MENMRSENFNQGVSMEGVKHAPEMANTNRRALRDIKNIIGAPHQHMAVSKRGLLDKP
 AAKNQSGHRPMTKFAATLANQPSIAPLPIGSERQKRTADSAFHGPADMECTKITSDD
 LPLPMSEMDEV MGSELKEIEMEDIEEAAPDIDSCDANN SLAVVEYVDEIYSFYRRSEG
 LSCVSPNYMLSQNDINEKMRGILIDWLVHYKLELLDETFLTVNIIDRFLARENVVRK
 KLQLVGV TAMLLACKYEEVSPVVEDLILICDRAYTRTDILEMERMITVNTLQFDM SVPT
 PYCFMRRFLKAAQSDKKLELMSFFIELSLVEYEMLKFPQPSMLAAAIYTAQCTINGFKS
 WNKCCELHTKYSEEQLMECSKMMVELHQKAGHGKLTGVHRKYSTFRYGC PAKSEPAV
 FLKSV AL*

Figure 7
 Page 14 of 32

ॐ नमो भगवते वासुदेवाय । नमः ॥

>20815 OsORF019753

ALLQKKDAELNTAKNSDLIKAHEEAIREAEKEISAALAERDKAIHDLQIAQSKYGEEIEA
RDLALADSDKKLKNVMAKLDLSLTSKFLSKKESWEKNVASVEESWRLKCESVKAESNG
HAGDELKKNLVELTVKHEKLKEEHDSFRDIADRMLEEKDREVAKLLENKDLHNSLEA
KAAVSINGNQNPQPAKQDRTR

MSFQDLEAGNARGLPRRGGGGRAGAAAAGAGASQAVASGVFQINTAVSTFQRLVNTL
GTPKDTPLRLRERIHKTRQHTQLVKDTSEKLKQASEADHRVEVSASKKIADAKLAKDFQ
AVLKEFQKAQRLAVEREAAYPFISQAGLPQSYNSSEVNNGADKLAEQRTALLESRRQE
LVFLDNEIVFNEAVIEERDQGIQEIQHQITEVNEIFKDLAVLVHDQGMIDDIDTHIENAVI
ATTQAKGQLSKAAKTQKSNSSLICLLLVIFGVLLIVIVLAA*

MAGNPAAAAPSSSGSSSVFLPPSPSDGELLRLPLHRLARDLSAVDTPAPFLRAAFASISRR
SKLLAAAFDDLLLCGAAGELPRASLCLREVLVLQRFKAIVADCSARSRMRLLESDE
MEAEELRELNHDLATLLDLLPVVELGLADDVLDVLALASRQCRCSPAPESEEALKASVL
SLIQEIEREIVPERERLEEILVEVGINDPASCSEEIESLEQEIGDRASEKWTASMIALVGLLR
YAKCVLFSATPRPSDSNSKADVEAEDGEPPVPPSPDFRCPISLDLMRDPVVVASGQTYDRE
SIDRWFSSGKSTCPKTGQVLANLELVSNKALKNLISKWCRENGVAMEACEASKSEQAQ
AVAANKAALEAARMTASFLVKKLSVSFSPDAANRVVHEIRLLSKSGSENRAFGVGEAGA
VPLLVPLLYSEDAGLQLNAV TALLNLSILEANKKRIMHADGAVEAVAHIMSSGATWRA
KENAAAAVLSLASVHSYRRRLGRNQSVVEKLVHLVRTGPTSTKKDALAALLTLAGERE
NVGKLV DAGVAEVALSAISKEETAAAVLAALAKRGGAEIVNIDGAVARLVAEMRRGT
DWARENATAALVLLCRRLGAPAVTQVMAVPGVEWAIWELMSIGTERARRKAASLGRI
CRRWAAASAADGERGGGCPVATVPPAMMAS*

MRRRRWRRRLAPVFRFYPTTEELICLYLRNKLDDGFRDDIERVIPVFDIYSVDPLQLSGTST
 RVYLLLPAAARRGSGWIFYCPRQEREVRGGRPSQTTPSGYWKAAGTTGVVYSAEPWPRA
 VGNKDRVEDDEREREGRWKMRSYRWVFPIAKRQITVLAEKCLPRDTEDELDGGGHVD
 DMTKLTLYLNEPGVLYNLKRRYALNEIYTYTGSILIAVNPFTRLPHLYNEYMMEQYKGVR
 LGELSPHVFAVADASYRAMVNDSSRSQSILVSGESGAGKTETTKFIMQYLYTVGGRAID
 DRTVEQQVLESNPLLEAFGNAKTVRNDNSSRFGKFVEMQFDANGRISGAARTYLL
 ERSRVVQINDPERNFHCFYQLCASGKDAELYKLGHGPSFHLYLNKSKTYELEGTNNEDEY
 WTKKRAMDIVGISRNDQDAIFRILAAILHLGNIEFSPGKEIDSSKIKDPTSNFHLQMAAKL
 FMCDPDLVSTLCTRAINTLEGAIKALDCSAAAANRDALAKTVYARLFDWLVENINKSI

Page 15 of 32

GQDVDSKVQIGILDIYGFESFKNNSFEQFCINFANEKLQQHFNEKPIGIIALLDEACMFPS
 THETFATKMFNRNFSSHRLEKTKFSETDFTISHYAGKVTYQTESFLEKNRDYTVAEHCNL
 LSSRCPLVSGLFGTLPEESLRSSYKFSSVASRFBKQQLQALMETLNSTEPHYVRCVKPNS
 VNQPQIFENQSVLHQLRCGGVLEAVRISLAGYPTRRTYAEFVDRFGVLVPELMLGSYDE
 RALTKGILEKMKLDNFQLGSTKVFLRAGQIAILDMRRAEVLNAARHIQGRFRTFITRKE
 FVKTREASISIQAYCRGCLARKKYMVKRETAATVQKYVRRWRLHRTYQQSHSAALLI
 QSCIRGFARHYFSVIKEQKAALVIQSLWRKRKVILFQQYRQATVAIQCAWRQKVARRE
 LRRLKMAANEAGALREAKNKLEKKLDDTLRLTLERRLRAAGEEAKSVEILKRDKLIES
 LSAKCAAASAAQSEHDKNLLLQRQLDDSLREITMLRSSKIMTAAERENSNLKNLVES
 LSKNNSLEYELTSARKGSDATMKKLKDVEGKCNHLQQNLDKLQEKLTNMENENHVL
 RQKALNMSPLNNMPMTTKAFPQKFATPIGLPNGEQKHGYETPPPAKYLASLPQSLTRSR
 RTRMPVERQEENHEILLRCIKENLGFKDGKPVTAIIYSCLLHWRAFESERTAIFDHVIEAI
 NNVLKGEEADGRLPYWLSNTSSLLCLLQKNLRSNGLFATPSRRSGGTGLGIGDKIVQTLRS
 PSKLMGRSDNLGQVDARYPAIFKQQLTACVEKIFGQLRDNLKKEISPLLSVICIAPKSS
 RAQPGKATKSPGIGAQPPSNHWDNIVKFLDLLMDTLHENYVPSFFIRKLITQLFSFINIQ
 LFNSLLLRECCFTSNGEYVKAGLSLLEKWTDTATDEFAGTSMHELNYRQAVGFLVIHQ
 KRKKKLEERNELCPNLSVRQIYRICSMYWDDKYNTQGISNEVVVSAMREEVNKDTQNL
 VSNSFLDDDLCPFSTEDLSIAIPAYVDIELPESLHHYASVQLLLKHHDPPQV*
 >23363 OsPN23363
 QRRRRRKRGRGRDGEGERGRGSGGSSGVAVTITTAPMTETEDDMAVAEEEEVAASA
 ETEEHVQRILLADAFTQVSEMLEAGRALFKNLAADFEDRLCSIHKERVERWEEIREL
 RARDAANEQARSLLHNAQLHLLATVRHDHT*
 >23390 OsPN23390
 TPTTSLMSIPPELAGAIPLIDRFQVEGFLKAMQKQIHSAGKRGFFSKKSVGPHVREKFTLE
 DMLCFQKDPIPTSLKISSDLVSRISILFHVILKYMIGIDSPAIIISLDERIELVAKLYKHTLKR
 SELRDELFAQISKQTRNNPDRAWLIRAWELMYLCASSMPPSKDIGAYLSEYVHYIAHGA
 TTDSVVRVLALNTLNALKRSVKAGPRVTIPAREEIEALLSSRKLTTIVFFLDETFEETIDYD
 MATTVADAVEELAGIILSVYSSFSLFECRKVVNGSKSSDVGNEEYIGLDDNKYIGDILLS
 EFKAADRNKGEILHCKLVFKKRLFRESDEAITDPMFVQLSYVQLQHDYILGNYPVGRD
 DAAQLSALQILVEIGFVDNPESCVEWISLLERFLPRQVAITRAKRDWELDIVSRYQLMEH
 LSKDDARQQFLRLRLTLPYGNSVFFSVRKIDDPIGLLPGRILGINKRGVHFFRPVPKEYLH
 SAELRDIMQFGSSNTAVFFKMRVAGVLHIFQFETKQGEEICVALQTHINDVMLRRYSKA
 RSATS AVSQNDVSQTYKPPNIEIYEKRVQELSKAVEESERKADLLNEELQKKTQKQERDM
 QKELEGLRDTLQSERQSIKEVTNDLDKLKSLCDEKDSLQASLMEKTRLETRLKSGGQQ
 ESSNRTGVSGNHFERDTLPTVGTVNNSIEMLAKEEELKSCKKELDASKELSKKLTMEN
 NLLDQKVORLERAKSEEKSNMERVYEDECCKLKSRIAELEQKLESRTSLNVTESTLAL
 RNAEVDTLQNSLDELREFKADVDRKNQQTAEILKRQGAQLIELENLYKQEQVLRK
 RYYNTIEDMKGKIRVFCRLRPLNDKELIEKDKNIVCSPDEFTVAHPWKDDKSKQHIYDR
 VFDANTTQEEVFEDTKYLVQSAVDGYNVCIFAYGQTGSGKTFTIYGSNNPGLTPRATS
 ELFRVIKRDGHKYSFSLKAYMVELYQDNLVDLLAKNATHQKLEIKKDSKGVVTVENV
 TVVNISFEELRAILRGSERRHTAGTNMNVESSRSHLISIIESTNLQTQSYARGKLSFVD
 LAGSERVKKSGSAGKQLKEAQSINKSLSALADVIGALSSDGQHIPYRNHKLTMMSDSL
 GGNAKTLMFVNVSPAESNLEETYNLSMYASRVRIVNDTSKHVAPKEIMRLKKLIAYW
 KEQAGKRSEDDLEEIQEERTPKEKADNRLTS*
 >23416 OsATPF

Figure 7
Page 16 of 32

MKNVTHSFVFLAHWPSAGSFGLNTDILATNLINLTVVVGVLIIYFGKGVLDLLDNRKQR
ILSTIRNSEELRRGTIEQLEKARIRLQKVELEADEYRMNGYSEIEREKANLINATSISLEQL
EKSKNETLYFEKQRAMNQVRQRVFQQAQVQALGTLNSCLNTELHFRITIRANISILGAME
*

>23484 OsPN23484

MDPIGDPSPSSRRSVKRRPPARSPELSPKAWGGEAPELIRRLLEELEEAAARLRGEKEAAE
EAARELQAELDAERASAETATSEAMLMIERLQREKAAAQMEARQFRRYAEGREDRERE
VQEELASLSDLAASYHSRLQSHGIDPDSFSDGEEEQHDEEDGEEVEQIDTAALQTDGSS
GGDSIGGMQVKAMVDDDEEEQFTPVEKEFEYTVDVRCASSTTKVSGAVVVGEFMGEG
NAAAGGLYARVEALEADRAAMRREIAALRAERAQLVMARAMARRLCREVVAEQKAK
KVAASPRSFVLGVCKWVLSIIFWRNRSSTAR YTFGLSTTFLGFLLLDRSTMLSPWRRL
PSPRQ*

>25358 OsAAK39589

MKEPAQVRTARGGAADGVEVGVEEEEEPPRSATVKQEEANAVLGAEGSRPFAMRELK
EDHEVAAGSGVKAASGERNGIGSADAQGSSYSQESMQQFSSHHDVAMDLINSVTGVDE
EGRSRQRILSFAAKRYISAIERNHDDPDAYYNWALVLQESADNVDPNSSSSKDALLEEA
CKKYAEATRLCPTLYDAYYNWAIADRAKMRGRTKEAEELWKQAILNYEKA VQLNW
NSPQALNNWGLGLQELSAIVPAREKQTIKTAISKFRAAIQLODFHRAIYNLGT VLYGLA
EDTMRSGKPGVSASEFYSSAIYVAAHALKPNYSVYRSALRLVRSMLPLPYLKVG YLI
APPENSAIAPHKEWERSQFVLNHEELQQVNASDQPPSQSPGHVDSGRKLFRI VVADIVSV
SACADLTLP PGAGLRIDTIHGPRFLVADNWETIDSWLDAIRLVYTI FARGRSDVLAGITG
*

>25381 OsAAK20062

MVINLEFIRAIVTADEILLDPLTIDVIPFVEQLTHHLPLKNLVCNGNQPGGDDHGEKHDD
SHGDQVPRLNEATGAEHELPEFQVLELALETVCSSFDVNVSGLERRATPVLEELTKNVS
TRNLDRVRTLKSDDLTRLLAHVQKVRDEIEHLLDDNEDMAHLYLTRKQLQNQQVEALIS
SAASNSIVPGGTSLRLNNSFRSSVSISATSMHLDNDVEDLEMLLEAYFMQLDGIRNRILS
VREYIDDTEDYVNIQLDNQRNELIQLQLTLTIASFGIAVNTFIAGAFAMNIQSKLYSIDDG
SFFWPFPVGGTSSGCFMICIVLLWYARWKLLGP*

>26210 OsAAK38489

MSSPLAVVSSFWDKDFLEKERGGGLDEQGLKIAENQETSQKNRRRLAESTRDFKKASSDD
KLSLFNSLLKSYQEEVDNLTAKRAFGENAFLNTYQKLYEAPDPYPALASMAEQDQKLSE
LETENRKMKLELEEYRAEAAHLKNQQATIRRLERNRQLEQQMEEKVREMVMKQRS
LAEDSQKTLEALKDRERALQDQLRQATESVKNMQKLHESAQSQLFELRTQSEEDRAAK
ETE VNLLLDEVERA QARLVSLEREKGDLSQLQTTNEDATNSSDYVDSSDILESSLNAKE
KIISLNAELRSIENTLSSERETHVNELKKLTALLSEKENALTELKKELQERPTRRLVDDL
KKKVQILQAVGYNSIEAEDWELATNGEEMSKLEALLDKNRKMEHELTQLKV KISEKS
NLLEEAEEKKIAELTAKAEEQQKLILKLEDDILKGYSSDTRRTSLNDWDLQEIGSNEVAE
GTDPRHAPQDQDQSSMLKVICNQDRFRTRLRETEELRLKEKYEMLVVELEKTKAD
NVQLYGKIRYVQDYSHEKIVSRGPKKYAEDVESGSSDVETKYKKMYEDDINPFAAFSK
KEKDQRYKELGLRDKITLSSGRFLLGNKYARTFIFYTIGLHLLVFTLLYRMSALS YLRR
LNTFSVDKNFPDMETGWMGNDRMRGRAFEPLCKLLYH*

>26688 OsPN26688

MASSSSSLGLGAIFQSGCPLPPRPAVRRAPTRRRRAVATKISCIGWDPEGVLGPPQGGHIV
RLEFRRLRLERDSAREAFAERQVREEHERRRQEREARVIPD TDAGLVEFFLDTEAREIEVEI
GRLRPRLNQPFDFYIQREIAQIKFSITRTAEMEDRLIELEAMQKVLLLEGVEAYDKLQNDL

Figure 7
Page 17 of 32

VSAKERLTKILQSSDKKSTLLEMVERNELNMSILTLLDENIASAKTNNQEEAVAFMENV
RSSILKYITV*

>29882 OsPN29882

HASANEHILAMEKEVENLQAQLKQESLLRQQEQQKLSEESLLRQQEQQKLGTGEQSHAA
SLVAEKKDLEEKIAALTKKASDEASEFAARKAFSMEDREKLESQ LHD MALMVERLEGS
RQKLLMEIDSQSSEIEKLFEENSALSTSYQEA VAVTMQWENQVKDCLKQNEELRSHLEK
LRLEQATLLKTSNTTIQPDGQNETSISFPPEFVTENLSLKDQLIKEQSRSEGLSAEIMKLSA
ELRKAVQAQNNLARLYRPVLRGIESNLMKMKQETYATIQ*

>29942 OsPN29942

EHESDEEKNEVDTSNEIKEIVIDDMEPKVFQAGFLFMYRDNLVGDDELSASSSDCSIFDT
LAGKLLAAADRYELPRLRLLCESYLCKHISVNSVATTLALADRHAMELKSVCCLKFAAE
NLSAVIRTDGFDYLDKDNCPALQSEILRTVAGCEEECSSGGKSQSVWGQLSDGGDTSGRR
VRPRV*

>29956 OsPN29956

FEIEWELIDEKKEELQKEAIRIAEERRATEYLKNESDIKQEKDNLRVQFKSNSETLSREH
KEFMSKMQQEHASWLSKIQQERQDLKRDIDIORVELLSAKARQMEIDSYLREREEFE
QKKAKELEHINSQKEMINTKLEHVAVELQKLKDERKEATLERERREQELSEIKGTIEALN
NQREKLQEQRKLLHSDREAITVQIQQLNVLEELKIDSENKQLSLLQHDKSKLGSDINVKD
NHHDNSHSSPKQRFGRKLDLSPVSTPISWVRKCAQVIFKRSPEKSASHDQFVQNGVPPK
VGDSVDVEDVNLDFAKVGQKRLNHLVSCDQTEVLEPKRKHRRSTIQKVNGGEITSNCL
SALEEKCSKNEHDEA

>29957 OsPN29957

MAKSSADDAELRRACAQAVAASGARGEVVSFSIRVAKGRGIFEKLGR LAKPRVLALT
KQSTKGEEAAKAFLRVLYSSGAVLEPAKLYKLKHLTKVEVISNDPSGCTFVLGFDNLS
QSVAPPQWTMRNIDDRNLLFSILTMCKEILSYLPKVVGIDFVELALWAKENTVTLDNQ
SSTQDGQEKSVTTQTERKVTVTENDLGSQAKDEEEDMEALLDTYVMGIGEADAFSER
LKQELVALEAANVYQLLQSEPLIDEVLQGLDAASATVDDMDEWLRFNMKLRHMRDI
ASIESRNNGLEMQSVNNKGLVEELEKLLDRLRIPQE

>29958 OsPN29958

SLQNEVSALEKQTLSLANDCLQSNKLRMEENALSTQVLKTNMRSSGDQNTVRTVKDME
LQKLHGTIKALQKVVTDTAVLLDQERLDFNANLQEARQIEVLKLKEILDDDLIEMNYE
QMLKDIQLDLIQISSGNKTGSLGQANKTVAQANEKMLDSHGIVGASSSHVRNDLRPPQS
ESFERDNYKRPPSELMVVKELSIDKQELPRSITEPHQEWKNKVIERLASDAQRLNALQS
SIQELKTNTEASEGLELESVR YQIREAEGFITQLIDSNGKLSKKAEEFTSEDGLDGDNDL
SRHQRKIM

>29961 OsPN29961

DSAMDHKYGGQKDGAPDEGSVGVPGRKTKETVTAADAIDALDTAEFEVKRLDQHQEG
QNNNGTTFQPNVIMQGGQSPSDYVLNVVSNVRPNDLEQALLSLPFSADALKIMSYLKEWS
MVPLKVELVCRVCLVLLQTHHSQLTTPSARSILTELKGLYSRVKECKDAIGFNLAAMD
HIKELLAMRSDAPFRDARAKLMEIRQEQSRRSDRSDGAEKRKKKKRRPSGES*

>29965 OsPN29965

LLEESTYAKGLASAAGVELKALSEEVTKLMNQNEKLASELASVRSPTPRRANGLRGTR
RDSISRRHEPAPRRDNNAGYEREKALEAVLMEKEQKEAELQRRIEESKQKEAFLESELA
NMWVLVAKLKSQGHDLDFDTKYIGS*

>29966 OsPN29966

VTDSVMKGTISAVEQESARQIASKDAEIAFLNEKLHQFRNSGLSLSEGRDKLYEETYNLR
 QQLVTLKSKLLNSEWGLSVSHYNNFEGAEDSKHRGNEKSSKDGITKENGSKGSNEDIFI
 DPTVLKHMMDRDELVAHFNMNMNQMKRQHDSTLQEKTEEIFRLKRENLKKEGPNPWHL
 RNNKEFELMRKKIWEVITKLDEVLVENKRTIRIKSDVFPQQD

>29967 OsPN29967

SPLLSSSPSTPRAPHPPTRRALTTLTLTLLSLAMSEPASPPPAMPEDAAPPQPQPEPAVP
 AGEAAAPSLERKEELLPVEEKISELDESQSQLMGRRLRGLKEDLLNWRSSLDTQVTKYKS
 ELSDIKTALNSEIEQLRSDFQELRTTLKKQQEDVSNLKNLGLQDATDNDGNKGS

>29968 OsPN29968

MEKSPLAWFRLLVNNEDVVAIKQMQLHLILGRLQDSNAVLTHFNEYSEQCFAEVSNDFA
 GKTRLLKSMKADLDHIFLKLGRGMSRLAATYPDAFPTGAMAETMDQRPDLESPLD*

>29969 OsPN29969

MEKSPPETAAAAAEVEARFRSLVDTGDIGAIRQTQHLILGRLQDSNAVLTHFNEYSEQCF
 AEVSNDFAKTRLLKSMKDDLDHIFLKLRSMSRLAATYPDAFPDGAMAKTMDQRPDL
 ESPLD*

>29970 OsPN29970

SSSDQAAEFTDMEGESSAVTSPFPALTSTTPNELEMTNKNNSNVVGGMTHSNSMPTLTAA
 KDGN TKVLPFEFRALEV CLESACRSLEEETSTLEQEAYPALDELTSKISTLNLERVRQIKS
 RLVAISGRVQKVRDELEHLLDDEMDMAEMYLTEKLTRQ

>30848 OsPN30848

PRVRPWNFELFKMPRRTDNAASANSVEPEKSEECLEFDDDEBEEVEEEEIEYEEIEEIEE
 EEEVEEDEDVVEEVEEVEDEEEDDEEEESDETEGVSKTKGVHQKDVTEKGKHAELLALPP
 HGSEVYVGGISSDVSSDLKRLCEPVGEVVEVRMMRGKDDSRGYAFVNFRTKGLALKA
 VKELNNAKLKGKRIRVSSSQAKNKLFIGNVPHSWTDDDFRKVVVEEVGPGVLKADLMK
 VSSANRNRGYGFVEYYNHACA EYARQEMSSPTFKLDSNAPT VSWADPKNND SASTSQV
 KSVYVKNLPKNVTQAQLKRLFEHHGEIEKVVLPPSRGGHDNRYGFVHFKDRSMAMRA
 LQNTERYELDGGVLD CSLAKPPAADKKDDR VPLSSNGAPLLPSYPPLGYGIMSVPGAY
 GAAPASTAQPMLYAPRAPPGAAMVPMMLPDGRLVYVVQQPGGQLPLASPPPQQAGHR
 SGSGGRHGG

>30854 OsPN30854

MQGEVDQPMQMVLRVKHPSSLGGGGGGGEEAGEASSRSALSVFKAKEEQIERKKME
 VREKVFAQLGRVEEESKRLAFIRQELEGMA DPTRKEVEVIRKRIDV VNRQLKPLGKTCV
 KKEKEYKEILEAYNEKNKEKALLVNR LIELVSESERMRMKKLEELNKTVDLSLY*

>30899 OsPN30899

VLLDSLKRKTYDDEL RREELLYFRRFQSASQKKGGSGIFRQGFSPSEG VDEGPYGLSRR
 IACKKCGDFHLWYTGRAKSQARWCQDCNDFHQA KDG DGWVEQSFQPVLFGLLHKPE
 LPHAYVCAESIIFDVTEWFTCQGMRC PANTHKPSFHVNASLLKQNSGKGSTSAQRGGGI
 PNGVNM DGGIDE EEF EWLQNALQSGMFESFGAQNEPPSPGSGSNAKGSN SSS

>19695 OsRACD

MSASRFIKCVTVGDGAVGKTCMLISYTSNTFPTDYVPTVFDNFSANVVVDGSTVNLGL
 WDTAGQEDYNRLRPLSYRGADVFLLA FSLISKASYENVSKKWIPELRHYAPGVPIILVGT
 KLDLRDDKQFFVDHPGAVPISTAQGEELRKLIGAAAYIECSSKTQQNIKA VFDAAIKVVL
 QPPKQKKKKKKAQKGCAL*

>19758 OsE2F1

MDSVVRTHIGHMLKMIHFCSMFLITFSYKIQA ILLSSQQKRKAPEESDVAESSDCMITSP
 GFAVSPMLTPVSGKAVKTSKSKTKNNKAGPQTPTSNVGSPLNPPTPVGTCRYDSSLGLL

Figure 7
 Page 19 of 32

TKKFINLLKQAPDGLDLNNAEAETLEVQKRRIYDITNVLEGIGLIEKTLKNRIRWKGLDDS
 GVELDNGLSALQAEVENLSLKEQALDERISDMREKLRGLTEDENNQRWL YVTEDDIKG
 LPCFQNETLIAIKAPHGTTLEVPDPDEAGDYLRRYRIVLRSTMGPIDVYLVSQFDEKFE
 DLGGGATPSGHANVPKHQPTVEFNTTNAGVGQCSNSVAVDNNIQHSQTIPQDPSASHDF
 GGMTRIIPSDIDTDADYWLISEGDVSITDMWKTA PDVQWDES LDTDVFLSEDV RTPSSH
 NQQPSAVGGPQM QVSDMHKP*

>20231 OsMADS45

MGRGRVELKRIENKINRQVTFARRNGLLKAYELSVLCDAEVALIIFSNRGKLYEFCST
 QSMKTLEKYQKCSYAGPETAVQNRESEQLKASRNEYLLKARVENLQRTQRNLLGED
 LDSLGIKELESLEKQLDSSLKHVRTTRTKHLVDQLTELQRKEQMVSEANRCLRRKLEES
 NHVRGQQVWEQGCNLIYERQPEVQQPLHGGNGFFHPLDAAGEPTLQIGYPAEHHEAM
 NSACMNTYMPWLP*

>21003 OsE2F2(367)

MTEAVLIRTP LKRKHDRENEAAESNDWMMSPGYTNPAGSPVPTPLSGKGSKAFAKSK
 AAKGQKSCPQTPLCASSPGNPVTPVGGCRYDSSLGLLTKKFLNLLKGAPGGIVDLNNAE
 ETLEVQKRRIYDITNVLEGIGLIEKKLKNRIRWKGIDDSRPGEVSDDMSILQADIEALSQ
 EHSVDDQQISEMRDKLRGLTEDENNQKWLYVTEDDIKSLPCFQNTLIAIKAPHGTTLEVP
 DPDEVNDYPQRRYRIVLRSTMGPIDVYLVSQFEEMSGMETPPRTVQPVSMDSLENPRT
 LAAEPNKA AESQPNIQDGLLMPSDAPSSSQDIGGMMKIVPSELDTADYWLSDAGVSI
 TDMWKTAPEVE*

>21044 Os018989-4003

MAPPCGDA AAAASAAPGLANLLIREGAGLPSRPERYPPFRPCTSDSFAPISREGDDIPPQK
 KSVSLRSGGGGNAAREEGGANRNGKKEKTGAQRITGWGLREFSKIVSKKVEAKGRIT
 YNEVADEIFAELKSITQNGLEFDEKNIRRRVYDAFNVLIAIRVIAKDKKEIKWMGLTNYR
 YEKIQKLEEVHKEKITRIKNKKKLLQEIEKQFDDLQNTILRNQASORPAESVNGILLPFLI
 KTSRKARVEIEISED SKFARFDENGAPFTMHDDVSILEAIRRNKGRAGLSIHP*

>20910 OsMADS14

MGRGKVQLKRIENKINRQVTFSKRRSKLLKANEISVLCDAEVALIIFSTKGKLYEYATD
 SCMDKILERYERYSYAEKVLISAESDTQGNWCHEYRKLKAKVETIQKCQKHLMGEDLE
 SLNLKELQQLEQQLENSLKHRSRKSQMLLESINELQRKEKSLQEENKVLQKELVEKQK
 VQKQQVQWDQTQPQTSSSSSSFMMREALPTTNISNYPAAAGERIEDVAAGQPQHERIGL
 PPWMLSHING*

>22824 OsPN22824

PASGDPVTFGWVCGAWRGVPAFQLRSNLKSPVRWPRGRRCHDLLIRKQFRFPLHEGS
 FSFRYLR SATFACIFHEPVFPALIVGRDRTRREDRDQGRPGGRTEQPAPAAGLASS
 RRSMPRSRGSELPQRASPRAPHLKTTACSDANGAHHRTVVDRSSPKLADRHSPRSPLH
 EKKRAGTRVAELETGLGKVQDELKKLREQ LATAEAAKKDAQVALEEAKKRVGTKGSP
 ASAAAASPRSPSPVAVETPKKTEQQLKVTEPPAE EESSINSPATDVFEVVRTESGDKENQ
 SAAGAEDGEVSCGIKAAALAEKELEEEETKKMIEEESIDATAIDSGEKENPEVAELKAKL
 MAKDMEVAILTADNAELKKQVGEAAEAAKKAEEEAASLVEQDLKERAAREARM
 GEQLRASEAARETLEAEMRRLRVQTEQWRKAAEAAPPVIGGDAH FVGHNGNGWGSPA
 TMPDDCDEGFGGKRKGAGIRMLGDLWKKKGSK*

>23367 OsAAG13527

MEGINGTVFAYGVTS SGTHTMHGDQNCPGIPLA IKDVFS LIQDVINDLLDPTGQNLRV
 REDAQGT YVEGIKEEVVLSPGHALS FIAAGEEHRHVGSNFNLFSSRSHTIFTLMIESSAH
 GDEYDGV MYSQLNLIDL AGSESSKTETTGLRRREGSYINKSLLTLGT VIGKLSEGRATHIP

Figure 7
 Page 20 of 32

YRDSKLTRLLQSSLSGHGHVSLICTITPASSNMEETHNTLKFASRAKRVEIYAARNRMID
 EKSLIKKYQREISSLKQELDQLRRGLIGGASQEEIMILRQQLLEEGQVKMQSRLEEEEEEAK
 AALMSRIQRLTKLILVSTKNNIPALTDTSSSHQRHNSVNEEDKVSTSQDSSMLVQNDSATK
 DSLSSASPDVDEINQLRCASGDHSSLAGSGPDEMGGITASDQMDLLIEQVKMLAGEIA
 FGTSSLKRLIEQSIEDPEGTKNQIDNLEREIREKRRHMRALQKLMESGEASVANASMM
 DMQQTITKLTAQCSEKAFELELRSADNRVLQEQLOQKNVEINELQEKVLRLEQQLTTNT
 EASPEQCTEHELHDLKSKLQLKEAESEKLYEHMKITEENRELVNQNSTLCEEVAYAKE
 LASSAAVELKNLAEEVTKLSVQNAKQAKELLIAQELAHSRVPGRKGRSAGRGRDEVGT
 WSLDLEDMMKLMELQARKQREAALEAAAEKEHLEEEYKKKFDEAKKKELSLENDLAGM
 WVLVAKLKRGA LGISDLNVDDRSINLADITNGTKENKADKNVAVVEKQLSDNTVKS LT
 ABEYRNPEFEPLLVR LKAKIQEMKEKETD SLGDKDGNSHVCKVCFESATAAVLLPCRHF
 CLCKPCSLACSECLCRTRIADRIITFT*

>26317 OsAAK72891

MEAVQPPVLSDR LNPLIHRS AFPASVVQPQDHGDSVPGLCSGSFIDTRGRLSSGSMTSE
 DSPALTPRWLSIKSNSSSDNCFEGSKRAVSWSDRHVFNPNGQVNYAEFMDLMEQELD
 QLDRLKGDVTGLENFALPDNGYIIGTHLGMSLDVMLIEIDERFNALKLLLATVFRKARE
 MDSSSVSDLQWEHELQLEVINITIGEFISGLQEEMERKLYEQISMTNSMSKNWQDAIAQF
 ASMRDDLGA LSKLLPSLQESHISHSKHETSSNR SNRWKYNIFGKKNKEDHSSRAEENK
 SFRKQKSMV VLEKSDFRHLNGMSKEEITYFKSEMSKLKRMHELDLQEKTEELFKFKRE
 KGLLTLKNDVEFEPLRKKIPQIISRMDDQIISKNIKMPSLCMTNDGLDERCISAKRIDSLYYE
 NQHLRGLLADNMKDVKELSSQLSEASKEMSIQLSSEDDLLRQIAKIKEEYEDLQIEAGVR
 DG VYQTITRKL LDDSMNSMHDAATNFSTELSSLEAMISEKEKALCLSNEENRMLKEKIA
 ELEQCLIQDKQEDPEVIKQESTEILRDIEVAPHISPRRSHETPKQDMQYDELVKLNSSLEI
 ASAALKEVENKNIDYNGIFTKNEQEKQLECILISIMKLSKEFVEIEQKLSVERSASRSEDLS
 DHCNHMVRQAVVLT KIGLWYKQMLETRRSELQKAEAKVVILGDKVNSHLNLLQKIYV
 TLD RYSP TLQQYPGLLDAFLKTCKLVAAFD*

>26539 OsPN26539

MVSGAAHSASTSGGGGGGSEGSPTGRAAPGMQGGGSAATPAASASASTPASETTVARRL
 DGLDIQGDDAPSSQPATSKKKKRGPGTRATGPDKGGRGLRQFSMKVCEKVESKGRTTY
 NEVADELVAEFADPNNNFASPDNDPNTPQFDEKNIRRRVYDALNVLMAMDIISKDKKE
 IQWKGLPRTSMDSVEELKTEIIGLKGRIDKKNAYLQELEDOFVGLQNLAQ RNEQLYGS G
 NAPSGGVALPFILVQTRPHATVEVEISEDMLVHFDFNSTPFELHDDSFVLKALGFSGKE
 PDDTQAWVGNGGECSTTPYHQSPQVARPNGVRLPTSPPIPGILKGRVKHEH*

>29946 OsPN29946

TMGTACKEGSHIGIIPRAMATLFDKIDKLKNQVEFQLRVSFIEILKEEVRDLLDPATAAVG
 KLENGNGHATKLSVPGKPPVQIREASNGVITLAGSTE VHVTTQKEMTACLEQGSLSRAT
 GSTNMNNQSSRSHAIFTITLEQMRKADPIMTLDGMP IEMNEDYLCAKLHLVDLAGSER
 AKRTGSDGLRFKEGVHINRGLLALGNVISALGDEKKRKEGAHVPPYRDSKLTRLLXDS PG
 GNSKTVM IACISPADINAEETLNTLKYANRARNIQNKPIVNRNPVADEMKRM RQQIEYL
 QAELVSARGGVVLD DVQGLRERISMLEQKNEDLCRELYDLRNHGYTDPCEPELQKIGT
 GYTKGEGLKRS LQSTEPFDVPM TDSVRAGSPKIDIDEVAKWEHTMLQDSMGKELNEL
 NRQLEQKESEM KMYGSDTVALKQHFGKKLLEEEEKRAVQQQERDRLLAEVESLNAD
 GQTHKLRDAQ LQKLKTLEAQ QILD LKKKQENQVQLLKEKQKSDEAAKKLQEEIHSIKA
 QKV VQLQH KIKQEA EQFRQWKATREKELLQQLRKEGRRNEYERHKLQALNQRQKLVL
 QRKTEEAAMATKRLKELLEARKSSGRDNS

>30852 OsPN30852

MARAGGHGMGNPNVGVIAVQADWENREFISNISLNVRLFDLLRFEATTKSKLASLNE
KLDILERKLEVLEVQVSSATTNPSVFN*

>31182 OsPN31182

MVSGVAHRPDDDGGRAASTFORPPQAGARPSLATPPPSGGAQSASTSGGSAGSPSSRS
EQHVPAAGMAAGAAAASTPISENTFLRLNDLDIHGDDAPSSQAPTSSKKKRGARAVG
PDKGGRGLRQFSMKVCEKVESKGRTTYNEVADELVAEFADPNNSILPPDPNPNAQQY
DEKNIRRRVYDALNVLMAMEIISKDKKEIQWKGLPRTSINDIEDLQTELVLKKSRIEKKN
TYLQELQDQFVGMQKLIQRNEQLYGSNIPSGGVALPFILVQTRPHATVEVEISEDMLV
HFDENSTPFELHDDSFVLKAMSSCGEEQIDGIHDLISNGGESSMPNIYRQQVQQPARST
NGTARLPSSPPPIGILKGRVKHEH*

>19697 OsTFX1

MSHSSDEDSEISDSEIDEYADKFYARLVAGEFKVKDGQSYSCPFCSGKKKKDFNINNLIQ
HASGVGAASNRQAKDKATHRALAKHLKNGLTKSSGQQSQTAVVEPQPLNRDEKFWW
PWMGVLVNVPTIEWKDGRQIGRSGNHLKEQLSRFCPLKIPLWNFRGHSGNAIVEFGKD
WHGFRNALAFEDYFGKEGYGRRDWKEKQNGSNLFGWVARAEDHTSPGLIGDHLRKN
GDLKTINDLENEGARKTDKLVANLANQIEVKNRHLQLEVTYNERTTSLEKMMGQREQ
LLQKYNEEIRKMQQLAQRHSQKIIDENQKLSELESKMSLNTRSKELDEIAAKSDYDRR
IIDQEKQKNAIKSSHLKLATLEQERADENVLKLVRHKKREKEAAVKKILKLEQQVDAKQ
KLELDIQQLKGKLEVMKHMFGDEDSALKNKIDELSEELQEKMDELDAMESLNQTLVIK
ERKSNTMQDARKELENGLDLLDGQSHIGIKRMGELDLEAFSKACRKMSSSEEDAETA
AILCSKWQAEIKNPDWHPFRFVLVDGQEKEIIEDDAKLQELKEEHGEDYRLVRDALCEI
NEYNPSSGRFPVGELWNFKDKRKATLKETVQFVLRQWRANRRKR*

>20080 Os005792-3529

MVQLLQLAIDCSAQHPDRRPSMSEVAARIDEIRRSSLGDRPATDSAGEGEEPSL*

>20257 OsCYCOS2

MENMRSENFNQGVSMGVKHAPEMANTNRRALRDIKNIIGAPHQHMAVSKRGLLDKP
AAKNQSGHRPMTKFAATLANQPSIAPLPIGSEKQRTADSFAHGPADMECTKITSDD
LPLPMMSEMDEVMSSELKEIEMEDIEEAAPDIDSCDANNSLA VVEYVDEIYSFYRRSEG
LSCVSPNYMLSQNDINEKMRGILIDWLVHYKLELLDETFLTNIIDRFLARENVVRK
KLQLVGVTA MLLACKYEEVSVPVVEDLILCDRAYTRTDILEMERMIVNTLQFDMSVPT
PYCFMRRLFLKAAQSDKKLELMSFFIIELSLVEYEMLKFPQPSMLAAAIYTAQCTINGFKS
WNKCCELHTKYSEEQLMECSKMMVELHQKAGHGKLTGVHRKYSTFRYGCAPKSEPAV
FLLKSVAL*

>20466 Os005750-3115

MADASSRTDTSTVLDTDKDNQMVDGQSGAIVPSNSSDRSDRSDKPMQKVLRLAQN
REAARKSRLRKAYVQQLESSKLKLASLEQEINKARQQGIYISSSGDQTHAMSGNGAMT
FDLEYARWLEEONKQINELRTAVNAHASDSLRLIVDGIMAHYDEIFRLKGVAADKADV
FHILSGMWKTPAERCFLWLGGFRSSELLKLLVNQLEPLTEQQLGLSNLQSSQQAEDA
LSQGMEALQQSLADTLAGSLGPSGSSGNVANYMGQMAMAMGKLTLENFLRQADNL
RQQTLHQMQRILITIRQAARALLAIHDYFSRLRALSSLWLARPRE*

>20534 Os018049-3655

MQQQQQHLHHQPSPPAGAASTWQLQQGAWHLRGSRFLPTQQLQEFCSLPVKSTTSP
SSASKATKPPQEEAASGGGSSSWTAPTQIQSMDAAELQRLKGKLYTMLEEVDRRYRRY
CEQMRALAAASFEAVAGERAAASYTRLASRTISRHFRSLRDGVVAQLQAVRKQLGEKDT
AVPGMTKGETPRLRVLDQCLRQHKA YQAGMLESHPWRPQRGLPERAVSILRAWLFEHF

LHPYPSDVDKHILARQTGLSRSQVANWFINARVRLWKPMVEEMYAEEMKDEEGSGQS
TQASNPGRV

>20559 OsHOS59

LDLFMTHYVLLLCSEFKEQLQQHVVRHAMEAVMACWELEQTLQSLTGASPREGSGATM
SDEEDNQVDSSESNMFDGNDGSDGMGFGPLMLTEGERSLVERVRQELKHELKQGYREK
LVDIREILRKRRAGKLPGDTASTLKAWWQAHSKWYPYPTTEEDKARLVQETGLQLKQIN
NWFINQRKRNNWHSNPASSSSDKSKRKRYRVVDF*

>20689 OsMYB

MGRQPCCDKVGLKKGPWTAEDQKLVSFLLGNGQCCWRAVPKLAGLLRCGKSCRLR
WTNYLRPDLKRGLLSETEEKTVIDLHEQLGNRWSKIASHLPGRTDNEIKNHNWNTHIKKK
LRKMGI DPVTHKPLY PAPPLADGGSPEQKVPEEEEEVEEKSSAAVESSTSTCAGHDVFCT
DEVPMHLHDDIVLPPPCDVVGDTAGSPAESSSTSTSSSGGGGIDEWLLPIMEWPESMYL
MGLDDVDMVTTAAPAMATSWEFEDPFNAYQRIALFDHHHELTWA*

>21036 Os003181-3684

MSTESGMLRGAGVGVVSGAVFSIEAVESCIEIWRSSSESGKYSIIFVLDISSLFSGRIVWEK
VSPALQRAVQSQMSLLSTPFIDNNDLFETGNTGGMSRDLINRIPKTTTSAATNPDPQETDN
CCAACLQDFGASQFVRVLPHCQHTFHARCIDNWLFRHASCPLCRAGVHIDHIHM*

>22896 OsAAD27557

MDSTKQDFQPRTFSIKLWPPSESTRMLVERMTKNLSTESIFSRKYGLLGKEEAHDNAK
RIEEVCFASADEHFKEEPPDGDGSSAVQLYAKETSKLMLEVLKRGPRTTVEPEVPVADTP
LEPADSVFDISGGKRAFIEADEAKELLSPLIKPGNAYKRICFSNRSFGIGAANVAGPIESI
KKQLTEVDISDFVAGRPEDEALDVMRIFSKALEGAVLRYLNISDNALGEKGVRAFEELL
KSQDNLEELYVMNDGISEEAAQALSELIPSTEKLKILHFHNNMTGDEGAMFIAEMVKRS
PNLESFRCSATRIGSDGGVALAEALGTCTRLKKLDLRDNLFGVEAGLALSKTLSKLPDL
VELYLSDLNLENKGTVAIINTLKQSAPQLEVLEMAGNEINAKASQALAECLTAMQSLKK
LTLAENELKDDGAVVIAKSLEDGHQDLKELDVSTNMLQRVGARCFQAIAANKPGFVQL
NINGNFISDEGIDEVKDILKSGENSVEVLGPLDENDPEGEAEDEEEEEEEEEEDDDQRRGL
RHRSPHPHPPPPLAITLQQIGNLAGGGGGGAGKRADTGDEREEEEEEEEGEDGGDAVAAR
RRDRPRDGHRRGIHRLRWEDKLPRLPQGGSPRHLQGRGLHTPLQHHQRHFYLCVSPG
ASCSLDLHIVRSLQYVDARVGPFDFTLVGGDTAEVSDIKFSNDGKSMMLTTTNNHIYVL
DAYGGDKRCGFSLESSPNVATEAAFTPDGQYVISDCLLEQSHRSYHCFEVGSPPSSNVCN
CVNCPNFLDSQQFQFELGA*

>23251 OsPN23251

MGGRKRALLVGINYPGTKAELKGCHNDVARMRRALVDRFGFDEADIRVLADADRSAP
QPTGANIRRELARLVGDARPGDFFHYSGHGTRLPAETGQDDDTGYDECIVPSDMNLI
TDQDFTLVQKVPDDCLFTTVSDSCHSGLLDKTKEQIGHSTKQNAQQIKREERSDSGT
GGFRSFLKETLKETVRDAFESRGVHIPHQSSRRNDDDEDEPHMGSSSHGGDRIKNRSLPL
STLIEMLEKETGKDDIDVGSIRMTLFSLFGGDASPKIKKFMKVMLTKLQEGQHGGVMGL
VGALAQEFMKAKLEGNEADALEPAMKQEVHSVHEAYAGTTARVSNGLISGCQTDQ
TSADATTPKGVSYGALSNAIQITLSEKSGRVTNKELVLRARELLSKQGYTQQPGLYCSD
KHTSVAFIC*

>23253 OsAAK00972

MATYYSSPGNERDSQAMYPADSGNSSYPVPSAIGNMLYPGNGSSGPYTEFSGIHQHQQN
FMELPGHPTAISQDSSSREPNNMVASYMDQRSFGPAKDMRNEMLMHLMDGAHNAGADL
IHNDTHSSAQIEFGLLNNHNSMSVAPAPGQGLSLSLNTHILAPSYPYWSAKTELLTPHSY
HGDDNRMKNMQSEASQAIRNSKYLKAAQELLDEVVSVWKSIIKQKAQKDQAEAGKSD

Figure 7
Page 23 of 32

NKEAEGGSKGEGVSSNPQESTANAAPEISAAEKQELQNKMAKLMAMLDEVDRKYKHY
YHQMQUIVSSFDMVAGSGAAKPYTAVALTQISKHFRLKDAINDQINVIRKKLGEESS
SGKEGKLTRLRYIDQQLRQQRAFFQYGLLQNAWRPQRGLPENSVSILRAWLFEHFLHP
YPKDSEKLMARQTGLTRSQISNWFNARVRLWKPMIEDMYKEEIGEADLDSNSSSDNV
PRSKDKIATSEDKEDLKSSMSQTYQPSQLGESKANIGMMSLGGAPAGFHNENQDDSF
MNLMLKDQRPGEAEGSLLHDAVAHHSDENARFMA YHLSGLGRYGNSNVSLTLGLQHP
DNRLSVQNTHQPGFAGAGEEINYNSTASLGVA AASSSDYESTNQIDQRQRFEPSPLMHDF
VA*

>23388 OsPN23388

MAAAATGGRCPMAPSPDWSSASPGYPKSEDSGEGGELRGWEASTAGACIPSAPTYAQH
SFAGRKFEHAVKHGAKNGLFVVTLGWL VDCVRRSMRLDESLYSIKNIGENGMPGFEFN
RLVGAPVSGNSCLPPMMFQEKTFSDTTEKHRLQTSRKEHDHDEFLEFNTDSIYIDPGISGE
MRKKYLNAATREGAKLLDHWFIGCHATYVVCEDASVKRYVGHSDNIVTPLWLKTAK
EKGLQRLVHLSSDLARQVATILENAQTFQENRKIGDVPSVNSNSSGVPSTQGEIDEIHQE
RQKFVEVAKKNVRDRRARMQSCVPIHPITPVKLMEICWTVSEPTTSACIYTEFSWSD
DAFEQQSTTFDANGDGKDDQSSDSFTRPLRESEKSEVIFKNHFLT VLFPIIDRFGLGPSS
RTFFSNGGFTRIQLDHIYNFYQENMSSDEINVALQTDSRHADRLRSLYASTESAERGFV
TFKRIDFLGSRSSFEGCLKRLSRENNSNVYELVIRA*

>23829 OsPN23829

MALSVEKTSSGREYKVKDLSQADFRLEIELAEVEMPGLMACRAEFGPSQPFKGARISG
SLHMTIQTAVLIETLTALGAEVRWCSCNIFSTQDHAAAIAARDSAAVFAWKGETLEEYV
WCTERCLDWGVGGGPDIVDDGGDATLLIHEGVKAEEFEKSGKVPDPESTDNAEFKIV
LTIIRDGLKSDPSKYRKMKERLVGVSEETTTGVKRLYQMQUETGALLFPAINVND SVTKS
KFDNLYGCRHSLPDGLMRATDVMIAGKVA VVCGYGDVGKGC AAALKQAGARVIVTEI
DPICALQALMEGLQVLTLEDVSEADIFVTTTGNKDIIIMVDHMRKMKNNAIVCNIGHFD
NEIDMLGLETYPGVKRITIKPQTD RWFVPETNTGIIVLAEGRLMNLGCATGHPSFVMSCS
FTNQVIAQLELWKEKSTGKYEKKVYVLPKHLDEKVAALHLGKLGARLT KLSKSQADYI
SVPVEGPYKPAHYRY*

>23830 OsPN23830

MGFLEDFAQSVEALPAMLQRNYSLMRELDKSLQGVQTGNEQRCQOEIEDIKHGLESISI
TYDPAKLKFSDEAIEEQKHCVRIDEKVALASQTYDLVDAHIQQLDQFMRKLEELRQEK
EAATTA AAAAAAAAAASVATGTPVAATVTASAGTSTADNTPKGRSSERGRGGRKKTA
KVPTEQPAPALDLELPVDPNEPTYCLCNQVS YGEMVACDNDCKIEWYHFGCVGVKEH
PKGKWYCPSCIGFQKKRKKGK*

>23832 OsBAB07943

MATFAKPENALKRAEELIHVGQKQAALQALHDLITSKRYRSWQKPLERIMMKYVELCV
DLRKGRFAKDGLIQYRIVCQQVNVSSLEEVIKHFQMQLSNEKAEQARNQAQALDLDV
EDLEADKRPEDMLSYVSGEKKGDRSDREHVTPWFKFLWETYRTVLEILRNNSKLEAL
YAMTAHKAQFQCKQYKRTTEFRRLCEIRNHLANLNKYRDQRDRPDLTAPESLQLYLDT
RVEQLKIATELSLWQEAFRSVEDIHGLMSMVKKTPKPSVLV VYYAKLTEIFWISDCHLY
HAYAWLKLFLYLQKSYNKNLSQKDLQLIASSVLLAALAVSPYDHKYGASHLELENEKDR
NLRIANLVNFSLDKRENREVPSRASLFSELAAGVVIACASQDVKDLYNLLEHDFLPLDL
VSKAQPLLSKISKIGGKLSSAPLVPEVFLSQYLPALAKLTTLRVLQQASQIFQSVKIDMLS
RMIPFFDFS VVEKISVDAVKHNFVAMKVNHLSGAVHFGKMDIESDCLSNHLSVLADSLN
KARSLIHPPVKKPSKLGENLTSLAA VVENEHKKRLARKSIEKREKEDLERQILEKEKEEEEK
KRLSVLKKS AEDERIRLLNDVKLREQERIRRLVEKEKIEAEELLQKQIKEIAKRGGKKP

Figure 7
Page 24 of 32

VLQGEVTKEAVMELAMNEQFKERQEMBEKKLQKTGKQMDYLERAKRQEEAPLIEQAFQ
KRLEVEKILHEQEQLREIELSKQHAGDLQEKNRLSRMLEHKNIFQERIVQRREAEF SRL
KKERDERTSQLISSRKRE RDTVRKLMYYLNL EEQRLQRLREEE EARKQEERRKREETER
KAKLDAIAAKQLQRERELEEKKEKQRMEALMGRGAGAAEPARTPD AAPVAQPAQPVA
APAAAAA AAAAAAAGKYVPKFKRGGDGGSSAGGQRP AVAPEQDRWGSRDD RPRPDMR
PLRQEAPPARDA APPARQDGPPGTWRPSRYSSSSSSSTWSSRN*
>24092 OsPN24092

MFKSARWRGCGGGAGGKAKAVFKLQFHATQVPEVGWEAMMVVTPRDAGRPTARTES
AQVADGACQWPAPVYEATKLPSSGKDKIYQFLVYDTGSTKAALLGEATLNLSEYADAF
KPWIVTLPLSGSPGAQLHVTIQRVVGCGGGGGGGGCGDDASENGGDVSPVVARTPPRKT
LQSQLSRCEDEEAEKARAAAAAADAMSPMQDGLVINKPPGMRFSAARNMRGPVDHLH
NSNSFDAISVSGSDGSSGRFTPKNNASMHSTFLQEGTNTLSPLRNTLTSSGDWSGSSAPD
ASTDGSTSNSGEAGLREADDVEKLRSEIATLTRKLDVSDMELQTLRKQIVKESRRGQD
LSKEVGSRLRDERDALRRECEALRGMKKTTHDANGSGKRLSSGEDPWSQIEELKQELGHE
KNLNGDLHLQLQKMQESNSELLAVKDLDEMLEQKNKEISLLHEETLEDPEAEYELAL
SNVHNAGHKIDISETSSVQEKEDMLDLAKTTDGIATSELQNKIVELSNEIELYKKDR
EDLEMQMEQLALDYEILKQENHDISSRLEQTQLREQLRMQYECSAHLSISDLEANVESL
ENELQEQSKRLEADIQEVMRAKVEQEQRRAKAEESLRKARWNNATTAERLQEEFKMLS
SQVSSAFSANEQLLMQARKBAEELQLQKGQLEELLQKAQEDLGSIQEQHRVKVQQLLT
LVDFKSKETDRLVMELKSKSDEFQNNQKRCNEAKLSVLSEEIDQLKAKIENLSNERDNLFE
ENEQKDKELAANCQKDMFLQDRDAEIALLNKELASIKDQVQTYLEEINTLKSSKNEKEE
MIERLQSEIRSLKFEYDNLKILMSTNDSEKHNLASQVLKLRRALLESREDVKQNGVKSD
DNHHATSKRIKHDDGTTGSRNVLPSTNRHNANGDCNGHRRDA AHDQSVKELEILKER
NTALEEELKELHGRYSEISLKFAEVEGERQQLVMTVRALKNSLR*

WNPDESSQATGGFTVAFSLSVVQSECQQLCEILRA'TPEAATADAAVQATARLANKDPV
KEKRDGSEGLSXAFRITDAAITAPNEGQGWRRNSTVPQEGYGTASVIPDQGIFLAASVYR
PVFEFMNKIGLMLPQKYWQLGNDGLLAFVNNFLKEHFLPAIFVDYRKCVQQAISSPA
RPRVHATSVYSPLVENGRPVLQGLLAVDIKEVLGWVQLMPNYATELVEYVRTFLERT
HERCRASYMEA VLEKQSYILLSRNDVESLMRLDPANLSLQNSFGQLDHSIPDAEAVEVEI
ELSDLLDMCPIKQENLIHDDQKLILLASLSLDSEYLADSVERLGESFINSSTMLENKNI
HQGRHTRSTS AIPKSLASLANEYRRLAIDCVRVLRLEMQLETTYHMQEMTKREYVEDQD
AEDPDDFIISLTTQIARRDEEMAPYIAESKRNYVFGGISSVAANASIKALAQMK SINLLGV
QQICRNSIALEQALAAIPSIDSEAVQQRIDRVRTFYELLNLPFESLLGFIAEHEYLFSAKEY
LSVLKVNVPGREMPMDAERRISOILGH*

PRVRREREKKDABEGKVEDVDEEKEEKEKKKKKKIKEVSHAWNVMNKQKPIWLRKP
TKEEYAAFYKSLTNDWEEHLAVKHFSVEGQLEFKAILFVPKRAPFDLFDTRKKQNNIKL
YVRRVFIMDNCEELIPEWLSFVKGIVDSEDLPLNISREMLQQNKILKVIRKNLVKKCVEL
FFEIAENKEDYNKFYEAFSKNLKLGIHEDSTNRTKIAELLYHSTKSGDELTSKDYVTR
MKEGQSEIYYTTGESKKAVENTSPFLEKLKKKGYEVLVMVDAIDEYAVGQLKEFEGKKL
VSATKEGLKLDESEDEKKRQEELKEKFEGLCVKIPEVLGDKVEKVVVSDRVVDS
VTGEYGWTANMERIMKAQALRDSSMAGYMSSKKTMEINPENAIMDELRRADADKN
DKSVKDLVMLLFETALLTSGFSLEDPNFTGTRIHRMLKLGLSIDEDESAEADADMPPLED
DAGESKMEEVD*

Figure 7
Page 25 of 32

MGRGKVELKRIENKISRQVTFARRNGLLKKAYELSLCDAEVALIIFSGRGRLEFESSSS
 CMYKTLERYRSCNYSQDAAAPENEINYOEYLLKTRVEFLQTTQRNLLGEDLGPLSMK
 BLEQLENQIEVSLKQIRSRKNQALLDQLFDLKSKEQQQLQDLNKDLRKKLQETSAENVLH
 MSWQDGGGHSSTVLADQPHHHQGLLHPHPDQGDHSLQIGYHHPHAAHHQA YMDH
 LSNEAADMVAAHPNEHIPSGWI*

>20072 Os000564-1102

MSAQAELSREENVYMAKLAEQAERYEEMVEFMEKVAKTVDSEELTVEERNLLSVAYK
 NVIGARRASWRISSIEQKEESRGNEDRVTLIKDYRGKIETELTKICDGLKLLSHLVPSST
 APESKV FYLKMKG DYRYLAEFKTGAERKDAAENTMVAYKAAQDIALAELPPTHPIRL
 GLALNFSVFYYEILNSPDRA CNLAKQAFDEAISELDTLSEESYKDSTLIMQLLRDNLTLW
 TSNISEDPPKIREVPKRDSSEGQ*

>20231 OsMADS45

MGRGRVELKRIENKINRQVTFARRNGLLKKAYELSVLCDAEVALIIFSNRGKLYEFCST
 QSMTKTLEKYQKCSYAGPETA VQNRESEQLKASRNEYLLKARVENLQRTQRNLLGED
 LDSLGIKELESLEKQLDSSLKHVRTTRTKHLVDQLTELQRKEQMVSEANRCLRRKLEES
 NHVRGQQVWEQGCNLIGYERQPEVQQPLHGGNGFFHPLDAAGEPTLQIGYPAEHHEAM
 NSACMNTYMPWPWL*

>20232 OsRAP1B

MGRGKVQLKRIENKINRQVTFSKRRSGLLKKANEISVLCDAEVALIIFST
 KGKLYEYATDSCMDKILERYERYSYAEKVLISAESDTQGNWCHEYRKLKAKVETIQKC
 QKHLMGEDLESNLKELQQLEQQLENSLKHRSRKSQMLLESINELQRKEKSLQBEENKV
 LQKELVEKQKVQKQQQLQWDQTQPQTISSSSSSFMREALPTTNISNYPAAAGERIEDVPA
 GQPQHVRIGLPPWMLSHING*

>20233 OsMADS6

MGRGRVELKRIENKINRQVTFSKRRNGLLKKAYELSVLCDAEVALIIFSSRGKLYEFGSA
 GITKTLERYQHCCYNAQDSNNALSETQSWYHEMSKLKAKFEALQRTQRHLLGEDLGPL
 SVKELQQLEKQLECALSQARQRKTQLMMEQVEELRRKERQLGEINRQLKHKLEVEGST
 SNYRAMQQASWAQGA VVENGAA YVQPPHSAAMDSEPTLQIGYPHQFVPAEANTIORS
 TAPAGAENNFM LGWVL*

>20668 OsMADS13

MGRGRIEIKRIENTTSRQVTFCKRRNGLLKKAYELSVLCDAEVALIVFSSRGRLYEYSNN
 NNVKATIDRYKKAHACGSTSGAPLIEVNAQQYYQQESAKLRHQIQMLQNTNKHVLGD
 NVSNLSLKEKQLESRLKGIKIRARKNELLASEINYMAKREIELQNDNMDLRTKIAEE
 EQQLQQVTVARSAAMELQAAAAAQQQQQNPFAVAAAQLDMKCFPLNLFEAAAQVQ
 AVAAQRQQIPTELNLGYHHHLAIPGATAADAPPHF*

>20698 OsFDRMADS8

MVRGRTELKRIENPTSRQVTFCKRRNGLLKKAFELSVLCDAEVALIVFSPRGRLYEFASA
 PSLQKTIDRYKAYTKDHVNNTIQQDIQQVKDDTLGLAKKLEALDESRRKILGENLEGC
 SIEELRGLEMKLEKSLHNIRLKKTELLERQIAKLKEKERTLLKDNENLRGKHNRLEAAAL
 VANHMTTTTAPAAWPRDVPMTSSTAGAAMDVETDLYIGLPGTERSSNRSETG*

>20700 OsMADS3

MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVALIVFSSRGRLYEYAN
 NSVKSTVERYKKANS DTSNSGTVAEVNAQHYQQESSKLRRQQISSLQNANSRTIVGDSIN
 TMSLRDLKQVENRLEKGIKIRARKNELLYAEVEYMQKREVELQNDNMYLRSKVVEN
 ERGQQPLNMMGAASTSEYDHMVNNPYDSRNLQVNIMQQPQHYAHQLQPTTLQLGQQ
 PAFN*

Figure 7
 Page 26 of 32

>20770 OsMADS5

MGRGKVELKRIENKISRQVTFARRNGLLKKAYELSVLCDAEVALIIFSTRGRLFEFSTSS
CMYKTLERYRSCNYNLNSCEASAALETELSNEYLYLKLKTRVEFLQTTQRNLLGEDLVP
LSLKELEQLENQIEISLMNIRSSKNQQLLDQVFELKRKEQQQLQDANKDLKRKIQETSGEN
MLHISCQDVGPSGHASEANQEFLHHAICDPSLHIGYQAYMDHLNQ*

>20778 OsMADS8

MGRGRVELKRIENKINRQVTFARRNGLLKKAYELSVLCDAEVALIIFSNRGKLYEFC
GQSMTRTLERYQKFSYGGPDTAIQNKENELVQSSRNEYLYLKLKARVENLQRTQRNLLGE
DLGTLGIKELEQLEKQLDSSLRHIRSTRQHMLDQLTDLQRREQMLCEANKCLRRKLEE
SNQLHGQVWEHGATLLGYERQSPHAVQQVPPHGGNGFFHSLEAAAEPTLQIGFTPEQM
NNSCVTAFMPTWLP*

>20837 OsBAA81880

MARERREIKRIESAAARQVTFARRRGLFKKAEELSVLCDADVALIVFSSTGKLSHFASS
SMNEIDKYNTHSNNLGKAEQPSLDLNLEHSKYAHLNEQLAEASRLRQMRGEELEGLS
IDELQQLEKNLEAGLHRVMLTKDQQFMEQISELQRKSSQLAEENMQLRNQVSQISPAEK
QVVDTENFVTEGQSSSVMTALHSGSSQSQDNDGSDVSLKLGLPCGAWK*

>20842 OsMADS15

MGRGKVQLKRIENKINRQVTFARRNGLLKKAEISVLCDAEVAIVFSPKGKLYEYAT
DSRMDKILERYERYSYAEKALISAESSEGNWCHEYRKLKAKIETIQCHKHLMGEDHE
SLNLKELQQLEQQLESSLKHISRKSHLMLESISELQKKERSLQEEKALQKELVERQKN
VRGQQQVQWDQTVQQAQAQAQPAQTSSSSSSMLRDQALLPPQNICYPPVMMGER
NDAAAAA VAAQGGVQLRIGGLPPWMLSHLNA*

>20847 Os008339

EGEGRAATGKRERIAIRRIDNLAARQVTFARRRGLFKKAEELSILCDAEVGLVVFSA
TGLFQFASTSMEQIIDRYNSHKTQRAEPSQLDLQGEDSSTCARLKEELAETSRLRQMR
GEELHRLNVEQLQELESGLSVLKTSSKILDEIDGLERKRMQLIEENLRLKEQVS
RMSRMEEMQPGPDSEIVYEEGQSSSVTNASYPRPPPDNDYSSDTSKLGLHS*

>19766 OsFDRMADS6

MGRGKVQLKRIENTINRQVTFARRSGLLKKANEISVLCDAEVALIIFSTKGKLYKYATD
SCMDKILERYERYSYAEKVLISAESDTQGNWCHEYRKLKAKVETIQKCQKHLMGEDLE
SLNLKELQQLEQQLENSLKHISRKSQLMLESINELQRKEKSLQEEKVLQKELVEKQK
VQKQQVQWDQTPQTSSSSSSFMREALPTTNISNYPAAAGERIEDVAAGQPQHVRIGL
PPWMLSHING*

>19877 OsRP5

MKIIFVFALLAIVACNASARFDALSQSYRQYQLQSHLLLQQQVLSPCSEFVRQQHSIVAT
PFWQPATFQLINNQVMQQQCCQQLRLVAQQSHYQAISVQAIQQQLQVGVVYFDQ
TQAQAQALLALNLPICGIYPNYIAPRSIPTVGGVWY*

>20910 OsMADS14

MGRGKVQLKRIENKINRQVTFARRSKLLKKANEISVLCDAEVALIIFSTKGKLYEYATD
SCMDKILERYERYSYAEKVLISAESDTQGNWCHEYRKLKAKVETIQKCQKHLMGEDLE
SLNLKELQQLEQQLENSLKHISRKSQLMLESINELQRKEKSLQEEKVLQKELVEKQK
VQKQQVQWDQTPQTSSSSSSFMREALPTTNISNYPAAAGERIEDVAAGQPQHERIGL
PPWMLSHING*

>20912 OsMADS18

MGRGPVQLRRIENKINRQVTFARRNGLLKKAEISVLCDAADVALIVFSTKGKLYEFSSH
SSMEGILERYQRYSFDERAVLEPNTEDQENWGDEYGILKSKLDALQKSQRQLLGEQLDT

Figure 7
Page 27 of 32

LTTKELQQLEHQLEYSLKHIRSKKNQLLFESISLQKKEKSLKNQNNVLQKLMETEKEK
 NNAINTNREEQNGATPSTSSPTPVTAPDPIPTNNSQSQPRGSGESEAQPSPAQAGNSKL
 PPWMLRTSHT*

>20914 OsMADS17

MGRGRVELKRIENKINRQVTFSKRRNGLLKKAYELSVLCDAEVALIIFSSRGKLYEFGSA
 GINKTLEKYNSCCYNAQGSNSALAGGEHQSWYQEMSRLKTKLECLQRSQRHMLGEDL
 GPLSIKELQQLEKQLEYSLSQARQRKTQIMMEQVDDLRRKERQLGELNKQLKNKLEAE
 ADSSNCRSAIQDSWVHGTVVSGGRVLNAQPPPDIDCEPTLQIGYYQFVRPEAANPRNG
 GGGDQNNNFVMGWPL*

>21116 OsMADS7

LVGGSWRSIGMGRGRVELKRIENKINRKVTFKRRNGLLKKAYELSVLCDAEVALIIFSN
 RGKLYEFCSTQSMKTLEKYQKCSYAGPETA VQNRESEQLKASRNEYLKLKARVENLQ
 RTQRNLLGPDLDLGIKELESLEKQLDSSLKHVRTTRTKHLVDQLTELQRKEQMVSEAN
 RCLRRKLEESNHVRGQQVWEQGCNLIGYERQPEVQQPLHGGNGFFHPLDAAGEPTLQI
 GYPAEHHEAMNSACMNTYMPWPWL*

>22834 OsPN22834

HASDGRTQLASWARIAMDRGDHHHLQQHQFLMPPAPVVPVQLCMPAMMADEQYM
 DLGGGGAAAAPGRGGAGERKRRFTTEEQIRSLSMFHAHHAKLEPREKAELARELGLQP
 RQVAIWFAQNKRARWRSKQLEHDYAALRSKYDALHSRVESLKQEKALTVQLHELRL
 REREERSGNGGAATTAASSSSCNGSGSEEVDDDDDKRNAAGCLDLEPPESCVLGGAT
 CATPADVSVESDQCDDQLDYDEGLFPESFCATPELWEPWPLVEWNAVA*

>23495 OsPN23495

MSSGAASPPPPPPAAEDGRVTSHVDPFLVEALDNPRHRLMVLRMAMDIQKFMQNPQL
 HEFEFQHFTPSYLRCAAHRVAQHYGLETTVADSLVDGSVSKIVAKKTSESCLPVIALSEV
 PSKQARNEHEAAEKLKFVICPRPKAFQNGAGDAGAKNNAARTVEERIEEYNKARARIFN
 GSISDIEGTSDLGALSVARDEPINVEPPVDENKVNTMNSRSRVAVFKDTEKDRSDPDYDR
 NYKRYVRGPVHDF

NVSPGGFNFVVPQFMQYGVGFMQSANMSRNQPSVYFGQPDLSMGSSSGTAVYPQWPT
 PAMIYPHCYDNLGHMISQVPVYQSFNHG*

>28517 OsBAB56078

MGWRWHDDGDDGGRGLGDIPDLAGGGGGGDGERCATRRVVQSRCHTEEEVEPGRFVR
 KCEKTEQLLRDCVGSIDIEALEKGLFGSIGSFLLDAERMTNDFLKSFGVPSINERESSFDG
 QPTGRHIGGQAPAGRHIEGTAKDTKQNDYAEFSSKITDV*

>29949 OsPN29949

MGRGKIVIRRIDNSTSRQVTFSKRRNGLLKKAKELSILCDAEVGLVVFSSSTGXLYEFSST
 NMKTVIDRYTNAKEELLGGNATSEIKIWQREAASLRQQLHNLQESHKQLMGEELSLG
 VRDLQGLLENRLAISLRNMRKDNLLKSEIEELHVKGSLIHQENIELSRSLNVMSQQKLEL
 YNKLQACEQRGATDANESSSTPYSFRIQANMPPSLELSQSQQREGECSKTAAPELGLH
 LP*

>29971 OsPN29971

EKSDVQSTLEMELDRRSNDWSVKLAEFQSEEQRLRERVRELAEQNVSFQREVTLLESNR
 IDVSNKITSLELQNKQLNDELQKVKEHDTLLKSSVELNDNLTKTAEGY

>31165 OsPN31165

MAAAAAAALITAASTPFPLVSFRSRRDGHLSLSPRRPGAGRCRASAPTFQGGPAASYA
 REMERLSAKESLLAFRDAGGFESLVSGKTTGMQKIDVNERIVGLERLNPTTPRPTTSPFL
 EGRWNFEWFGDSSPGALAAARLLFERSPTTVAHFTGLDVLIKDGYSKISSNVKFLNTVQS

Figure 7
 Page 28 of 32

MASNAAAAAVSLDQAVAASAAAFSSRKQLRLPAAARGGMRVRVRARGRREAVVVAS
 ASSSSVAAPAAKAEIIVLQPIREISGAVQLPGSKSLNRILLLSALSEGTTVVDNLLNSED
 VHYMLEALKALGLSVEADKVAKRAVVVGCGGKFPVEKDAKEEVQLFLGNAGTAMRP
 LTAAVTAAGGNATYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTECPPVRVKGIG
 GLPGGKVKLSGSISSQYLSALLMAAPLALGDVEIEIDKLISIPYVEMTLRLMERFGVKA
 HSDSWDRFYIKGGQKYKSPGNA YVEGDASSASYFLAGAAITGGTVTVQGCGTTSLOGD
 VKFAEVLEMMGAKVTWTDTSVTVTGPPREPYGKKHLKAVDVNMNKMPPDVAMTLAV
 VALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNITAIT
 YDDHRMAMAFSLAACADVPTIRDPGCTRKTFPNYFDVLSTFVRN*

>22858 OsPN22858

ASFRTVGAKITQETGDFVSDAEGDPDKPTDGFSSIDEAIGALHEGKFVIAVDDESGDNE
 GDLVMAATLADPESIAFMIRNGSGIISVGMKEEDLTRLMPMMSPIAEIEDISAAASTVTV
 DARVGISTGVSAADRAKTIFTLASPDSPKPTDLRRPGHIFPLKYRNGGVLKRAHTEASVD
 LVALAGLRPVSVLSTVINPVDGSMAGMPVLKQMALEHDIPVSIADLRYYRKRREKLVEL
 IAVSRLPTKWGLFRAYCYQSKLDGTEHIAVAKGDIGDGEDVLVRVHSECLTGDI LGSAR
 CDCGNQLDLAMQLIDKAGRGVLVYLRGHEGRGIGLGQKL RAYNLQDDGHD TVQANVE
 LGLAVDSREYGIGAQLRDMGVRTMRLMTNNAKFVGLKGYGLAVVGRVPVISPITKE
 NQRYLETKRTKMGHVYGSDLPGNVPEEFLNPDDIAGDQDEDDTHN*

>22866 OsPN22866

PRVRRSGRFFFLSPPTPTPIDLHPESLLLTAGELPAAAEMATRYWIVSLPVQTPGSTANS
 LWARLQDSISRHSFDTPLYRFNVPDLRVGTLDSSLALSDDL VKSNVFIEGVSHKIRRIE
 LERAGGVESGALTVDGVPVDTYLTRFWDEGKYPTMSPLKEIAGSIQSQVSKIEDDMKV
 RGAEYNNVRSQNLAINRKQTGSLAVRDLNLVKPEDMVTSEHLVTLLAVVPQYSQKD
 WLSSYESLDTFVVPSSKKLYEDNEYALYTVTLFAKVVDNFKVRAREKGFQVRDFEYS
 SEAQESRKEELEKLMQDQEAMRASLLQWCYASYSEVFSSWMHFCLVRVFVESILRYGL
 PPSFLSAVLAPSQKGEKKVRSILRNSVGNVHSIYWKSEDDVG VAGLGGKAVFLE*

>22874 OsPN22874

GTNPGFRVGEIRLSNRDIYFGTLLGNTPEGSGRYVWSDGCTYDGEWRRGMRHGQGKT
 MWPSGATYEGEYSGGYTYGEGTYTGSDNIVYK

>23022 OsPN23022

KYAERGLRSLAVARQEVPEKSKESAGGPWQFVGLLPLFDPPRHDSAETIRKALHLGVNV
 KMITGDQLAIGKETGRR LGMGTNMYPSSALLGQNKDASLEALPVDELIEKADGFAGVFP
 EHKEYEIVKRLQEKKHIVGMTGDGVNDAPALKKADIGIAVADAIDAARSASDIVLTEPGL
 SVIISAVLT SRCIFQRMKNYTIYAVSITIRIVLGFLLI ALIWKYDFSPFMVLIIAILNDGTIMTI
 SKDRVKPSPLPDSWKLKEIFATGIVLGSYLALMTVIFFWAMHKTDFFTDKFGVRSIRN
 SEHEMMSALYLQVSIVSQALIFVTRSRWSFIERPGLLLVTAFLMLAQLVATFLAVYANW
 GFARIKGIGWGWAGVIWLYSIVFYFPLDIFKFFIRFVLSGRAWDNLL ENKIAFTTKKDYG
 REEREAQWATAQRTLHGLQPPEVASNTLFNDKSSYRELSEIAEQAKRRABEARLRELNTL
 KGHVESVVKLKGLDIDTIQQNYTV*

>23053 OsPN23053

SSFLWGYLVSPHGGALVDYYGGKRV MAYGVALWSLATFLSPWAAARSLWLFLSTRVL
 LGMAEGVALPSMNMVLRWFPRTERSSAVGIAMAGFQLGNTIGLLLSPIIMS RAGIFGPF
 VIFGLFGFLWVLVWISAISGTPGENAQISAHELDYITRGQKL VKTQSGGERLRKVPPFSKL
 LSKWPTWALISANAMHSWGYFVILSWMPVYFKTIYHVNLR EAAWFSALPWVMMAVL
 GYVAGVVS DRLIQNGTSITLTRKIMQTIGFVGPGVALLGLNAAKSPVIASAWLTIAVGLK

Figure 7
 Page 30 of 32

SFGHSGFLVNLQEIAPQYAGVLHGMSNTAGTFAAILGTVGAGFFVDRMGSSFRGFLILTSL
LYFSSTLFWDIFATGERVDFDGTG*

>23059 OsPN23059

MAASLQAAATLMQPAKLGGRASSAALPSRPSSHVARAFGVDTGAAGRITCSLQSDIREV
ANKCADAAGLAGFALATSALLVSGASAEQVPRRLTFDEIQSKTYMEVKGTGTANQCPT
VEGGVDSFAFKAGKYNMKKFCLEPTSFTVKAEGVAKNAPPEFQKTKLMTRLTYTLDEI
EGPLEVSSDGTIKFEEKDGIDYAAVTVQLPGGERVPFLFTIKNLVATGKPESFGGPFLVPS
YRGSSFLDPKGRGGSTGYDNAVALPAGGRGDEEELAKENVKNASSSTGNITLSVTKSKP
ETGEVIGVFESVQPSDSDLGAKVPKDVKIQGVWYAQLE*

>23061 OsPN23061

MAMATQASAAKCHLLAAWAPAKPRSSTLSMPTSRAPTSRLAAAEDQPAATAATEEKKP
APAGFVPPQLDPNTPSPIFGGSTGGLLRKAQVEEFYVITWTSPKEQVFEMPTGGAAIMRE
GPNLLKLARKEQCLALGTRLRSKYKINYQFYRVFPNGEVQYLHPKDGVPYPEKVNAGRQ
GVGQNFRSIGKNVSPIEVKFTGKNVFDI*

>23426 OsRBCL

MSPQTETKASVGFKAGVKDYKLTYYTPEYETKDDILAAFRVTPQPGVPPEEAGAAVA
AESSTGTWTTVWTDGLTSLDRYKGRCYHIEPVVGEDNQYIAYVAYPLDLFEEGSVTNM
FTSIVGNVFGFKALRALRLEDLRIPPTYSKTFQG
PPHGIQVERDKLNKYGRPLLGTIKPKLGLSAKNYGRACYECLRGGLDFTKDDENVNSQ
PFMRWRDRFVFCAEAIYKSQAETGEIKGHYLNATAGTCEEMIKRAVFARELGVPVIMH
DYLTGGFTANTSLAHYCRDNGLLLHHIRAMHAVIDRQKNHGMHFRVLAKALRMSGGD
HHHAGTVVGKLEGEREMTLGFVDLLRDDFIEKDRARGIFFTQDWVSMGPVIVASGGIH
VWHMPALTEIFGDDSVLQFGGGTLGHPWGNAPGAAANRVALEACVQARNEGRDLARE
GNEIIRSACKWSPELAAACEIWKAIKFEFEPVDKLD*

>29982 OsPN29982

APGARVFLARPLLRRSPRGVACALRRRPSKYKNKIQNEEVVVEDDIGGGGEDDDDALE
ALFKQLEEDLKNDDLSVEDDDDGISEEDMARFEQELAEAIGDIADADESSEGSSLGSEA
YGNDEKTDEIKRPELKNWQLKRLARALKIGRRKTSIKNLAGEGLDRTLVIELLRNPPPK
LLFMSDSLPEDEPSKPEIKEIEPSVVDNADVTTETKPTQLPVHVMCAEWSSQKRLKKV
QLETLEVVYSRTKRPTNTMISSIVQVTSLPRKTIVKWFEDRREQDGVDPHRVAFKRSLSE
TIAS

>30846 OsPN30846

APRRPSTFLNAVALGNVGAGKSAVLNSLIGHVPLPTGENGATRPIVVLDLQRPGLSSKS
IVLQIDSKSQRSASSLRHSLQDRLSKGASSGSSRGRVEGINLKLRTSTAPPLKLVLDLPGI
DQRAVDDPMINEYAGHNDAILLVVIPAMQAADVASSRALRLAKDIDADGTRTVGVISK
VDQAEQDAKTIACVQALLNKGPKNLPDIEWVALIGQSVAIASAQAAGSENSLETAWN
AEAETLRSILTGA PKSKLGRIALVDTIK

>30974 OsPN30974

MGALLSSPNSKNQPWEHGEASKADSSKKLRMSAPPLSGGYDHPGLIPGLPDEISLQILAR
MPRMGYLNAKMVSRWSKAAITGVELYRVRKELGVSEEWLYMLTKSDDGKL VWNADF
PVCQWQRLPLMPGISHGGECKRGIPGLWLGDLLSAGIRVSDVIRGWLGQRDSLDRLPF
CGCAIGTVNGCIYVLGGFSRGSAMKCVWRYDPFVNAWQEVSSMSTGRAFCASLLNN
KLYVVGVSCKGNGLAPLQSAEVFDPRTGIWVGALTLVSKGPSSTSCLLVELVKPIA
TGMTSLGGKLYVLQSLYSGPFFVDVGGEIFDPETNSWAEMPVGMGEGWPARQAGTKLS
AVIDGDLYALEPSTSFDRGKIKIYDPQEDAWKVAIGQVPVGDFAESECYPYLLAGFLGKLN

Figure 7
Page 31 of 32

LIKDVD SKINIMQTDVLKPVELSAPGNGPTCQNQQLFSEQETNLWKVIVSKNLAAAELV
SCQVLNI

Figure 8

>19695 OsRACD

atgagcgcgtctcgggtcatcaagtgcgtcaccgtgggggacggcgccgtgggcaagacctgcatgctcatctcctacacctccaacacct
tccccacggactatgttccaactgttttgataacttcagtgcaaatgttggtcgtatgggagcactgtgaacttgggggttgggatacagca
ggacaaggagactacaataggctacgcccgttgagctatcgtggcgctgatgttttctgctggccttttctctgacagcaaaagcaagctatg
agaatgtttctaaaaagtggatacctgaattaaggcattatgctcctggttgccaataattctcgttggaaacaaagcttgatctcggggatgat
aagcaattttcgtatgacacctggtgctgtacctatttccactgctcaggggcgaagagctgaggaaactcattggtgcagcggcatacattg
aatgcagttcaaaaacacagcaaaacatcaaggcagtttctgatgctgcgattaagggtgttctccagcctccaaagcaaaagaagaagaa
gaaaaaggcgcagaaaggatgtgccatctgttaa

>19758 OsE2F1

atggattctgtttaggacccacattggtcatatgcttaagatgatacatttttgttctatgttttcttacttttagcgactataaaatacaagctat
tctctctcacagcaaaaaagaaaagcaccagaagaaagtgtgtgctgagctagcgactgtatgattactagccctggatttgcgtcagct
ccaatgtcactccagctcttgggaaagctgtaagacttcaagtcaaaagactaagaacaacaaagctgggcctcagacacctacatcaaa
tgttgggtcaccacttaatccccaaactcctgttggtacatgccgttatgacagttcgttagggccttctcacaagaagtccataaatttgcgaag
caggctccggtggcatttctagattgaataatgctgcagaacactagaggttcaaaaacggcgatatatgatattaccaatgtacttgaag
gaatagggttgattgaaaagacactcaagaacagaatccgttggaaaggccttggtgattcaggagtggaattagataatggtctttcagcttt
gcaggcagaagtgaaaatcttagtctgaaggagcaagcattagatgagcgcataagtatatgcgtgaaaaactaagaggttaactgaag
atgaaaacaatcaaatggtcttatgtgacagaagatgatatcaagggtattaccctgcttccagaatgaaacactaattgcaataaaagctc
ctcatggcactacacitgaagtcacagatcctgacgaggctggtgattatctccagagaagatatagaatctattaagaagtacaatgggcc
caatagatgtctacttagttagcaattttagagaaatttgaggacttgggtgggggtgcaacaccatcagggcatgcaaatgtaccaaac
atcaacctactgaagtcttcaatacaacaaatgctgggggttgggcaatgtagcaactcagttgctgtggacaataattcagcacagccaga
cgattcctcaggatcctagtgttccatgatttggagggaatgacaaggattatccctcagatattgatactgatgctgattactggtcatatc
agaagggggatgtcagcattactgatatgtgaaaacagcaccagatgtgcagtgggatgagagcctggatacggatgtcttctctatctgaag
acgttagaacgccaaagttcacacaatcagcagccatctgcagtgggcgggccacagatgcaggtctcagacatgcataaaccttag

>20231 OsMADS45

atggggagggggtcgggtggagctgaagaggatcgagaacaagatcaaccggcaggtgacgttcgccaagcgcagggaatggcctgctc
aagaaggcgtacgagctctccgtcctctgcgacgccgaggtcgccttcatcatcttctccaaccgcggcaagctctacgagttctgcagca
cccagagcatgactaaacgcttgagaagtatcagaaatgcagttacgcaggacccgaaacagctgtcaaaaatagagaaagtgcgaat
tgaaagctagccgcaatgaatacctcaaactgaaggcaagggttgaaaattacaacggactcaagaaatttgcgtgggtgaagatcttgatt
cattaggcataaaagagctcgagagccttagagaagcagcttgattcatccctgaagcacgtcagaactacaaggacaaaacatctggttga
ccaactgacggagcttcagagaaggaacaaatggtttctgaagcaaatagatgccttaggagaaaactggaggaaagcaaccatgttcg
cgggcagcaagtgtgggagcagggctgcaacttaattggctatgaacgtcagcctgaagtgcagcagcctcttcacggcggaatgggtt
cttccatccacttgatgctgctggtgaaccacccttcagattgggtaccctgcagagcatcatgagg
cgatgaacagtgctgcatgaacacctacatgccccatggctaccatga

>21003 OsE2F2(367)

atgacggaggccgtgctcatccggacgccattaaagcggaacatgacagagaagagaatgaagctgctgaatcaaatgactggatgatg
agccctggatacactaatccagcaggcagcccagttccaacaccgctttcaggaaaaggttcaaaagcttttccaaatcaaaagctgcaaa
aggccagaaatcttgcctccagaccctttgtgcgtagtctcaggcaatccggttacacctgttggtggtgctgccgatgatagctcccta
gggctattgacaaaaaagttcctaaacttgctaaagggtgcacctggcgcatagttgattgaataatgccgcagaaacactggagggtaca
aaaaaggcgtatatatgacatcactaatgtccttgaagggtatgggctgatagaaaagaagctcaagaacaatatccgttggaaagggaattg
atgactcccgaccaggagaagttagtgtatgtccatcttacaggctgatattgaagccctctcactgcaggagcacagcgtagatcaa
caataagtgaatgcgagataagttgaaggactcacagaagatgaaaataaccaaaagtggctatatgttactgaagatgacatcaagtc
ttgccctgttccagaatcagacactgatcgcaatcaaaagcgctcatggtacaacttggagggtcccagatcctgatgaagtgaatgattat
ccacagaggagatataggattgttctaaggagtactatgggtccaatagatgtctacctagttagtcaatttggaggagatgagtggtcag

Figure 8
Page 1 of 14

actcctccaaggactgtgcagccagtaagcatggattctctagagaatcccaggacgccattggctgcagaaccaacaagctgcagagt
cacagccaaatattcaagatgggctgctaatacctctctgatgctccttctagttcacaggatattggcgggatgatgaagattgtcccttcagaa
cttgataccgatgcagactactggctcttatcagatgccggggttagcattaccgacatgtggaagacagcaccagaggtggagtga
>21044 Os018989-4003

atggcgctccctgcggcgatgccggcggtgcctccgccgcgcccggcctggccaacctgtcattcgcgaaggcgccggcctcc
cttcgctgccgaacggtagccgccattccggcctgtacttctgattcctttgctccaatctctaggggaaggggacgatattccccccaaaa
aaaatcggtttctttacgcagcggcgcgcggaatgcggcagagagggaggaggcggtgcgaacaggaacgggaagaaggaga
agaccggggcgagaggaatcaccgggtgggggctccgtgagttcagcaagatagtttctaagaaagttgaggccaaaggaagaaccac
atataatgaggttgccgatgagattttgcggagctgaagtcattacgcagaacggtctggagttgatgagaagaatattaggcgagggt
atatgatgctttcaatgtgctcattgcaattcgtgttattgcaaaagataaaaaggagataaagtggatgggccttactaattatagatacga
gatacagaagttggagggaagttcacaagaactcatcaccaggatcaagaataagaagaagcttctcaggaaattgaaaagcagttgat
gaccttcagaatattacattacgcaaccaggctagtcagaggccagcagaaagtgttaatggcatcctccttccgttcttattgatcaagacat
cccgaagcaaggggtggaaattgagatttcggaagattcgaaattgcacgggttcgactcaacgggtgcaccattcaccatgcatgatgat
gtatcaatcctgaagccatcaggcgtaacaaaggaagactggcctctccattcacccttaa
>20910 OsMADS14

atggggcggggcaaggtgcagctgaagcggatcgagaacaagatcaaccggcaggtgaccttctccaagcgcaggtcaaaactgtc
agaaggcgaatgagatctcgtgctctgcgacgccgaggtcgcgctcatcttctccaccaagggcaagctctacgagtacgccaccg
actcatgtatggacaaaatcctgaacgttatgagcgtactcctatgcagaaaaggtccttatttcagctgaatctgacactcagggcaactg
gtgccacgaatataggaaactgaaggtaaggttgagacaatacagaaatgtcaaaagcacctcatgggagaggatctgaatcttgaatc
tcaagagctgcagcagctggagcagcagctggaaaattcgttgaacatatcagatccagaaagagccaactaatgctcgagtccattaa
cgagcttcaacggaaggaaaagtcactgcaggaggagaataaggtctacagaaagactggaggagaagcagaaagtcagaaagcaa
caagtgaatgggaccagacacaacctcaacaagttcctcatcctccttcatgatagggaagcccttccaacaactaatatcagtaac
tacctgcagcagctggcgaaggatagaggatgtagcagcagggcagccacagcatgaacgcattgggctgccaccatggatgctga
gccacatcaacggctaa
>22824 OsPN22824

cccgcgtccggagaccccgtaacgttcgggtgggtgtgcgggtgcgtggcgtggcgttcccaaggcgttccagctgcgttcaaatgtgaat
caccggtgcgggtggcctctgtgggagggcgggtgccacgatttattaatagcgaacaattccgttttcccttgcacgagggctccttttcttgc
cgctacctgcgcagcggcacattcgcctgtcattttccatgagccagtttccccgcgttctctcctcgttagggcgacccagcgacgcg
gcgagaggatagagaccagggggcgccggggcgacggacagagcaaccagcgccgggtgccgggttagctagctcgcggcgagc
atgccgagatccaggggtccgagctgccgcagcagcgtcggcgggcgccgtgcacctcaagacgacggcctgctcgatgcg
aatggcgcgaccaccgtacagctgctgaccggagcagcccaagctcggcaccgacctcggcgcgagcccgtgcatgagaag
aagcgcgggggacgcgggtgggtgagctggagacgaagctgggcaaggtgcaggacgagctcaagaagctgcgggagcagctc
ccaccggcgaggccgaagaaggacgcgaggtcgcgctcagaggagctaagaagcgcgttggaccaaggggagccccgcttc
ggcgccgggctctccgcgtccccatctcgggttgcgggtcgaacccccagaagactgagcagcagctgaaggtgacggagccgc
cgccgaggaggagagcagcataaactccccgcgacggatgtgttcgaggtcgtccggaccgaatcgggggataaggagaaccaga
gcgcgggcgggcgccgagggaggtgagctgtggcataaaggccgttggctgagaaggaactggaggagggaagagaca
aagaagatgatcgaaggagagcattgacgcaacggccattgactccggcgagaaggagaacccggaggtcgcggaactaaaggcc
aagctgatggcgaaggacatggaggtcgcatactacggcgacaatgccgagctcaagaagcaggttggtaagcccgaggaggcg
gccaagaaagctgaggaggaggcgggcgccaaggcgtccttggctgagcaggacctgaaggagagagcgggcggggaggcccgca
tggcgagcagctgaggcgctcggaagccgctcgggagacgctggaggccgagatcgggcgctgcgctgcagaccgagcaatgg
cgaaaggccgggagggcgccccccgggtgattggcggggatgccacttcgtcggccacaatggcaacggctggggctccccgcg
acgatccgggacgactgcgacgacgaggggttcggtggcaagaggaaggagccggtatccggatgcttggcgacctgtggaagaag
aaggggagcaagtga
>23367 OsAAG13527

atggagggtataaatggtacggtttttgcttatggtgtcacaagtagtgg
gaagaccacacaatgcatggtgatcagaattgccttgaataatccat

Figure 8
Page 2 of 14

tagccatcaaggatgtattcagcttaattcaagaigtataaatgatctacttgccaaaatctgctgtgagggagatgcaca
 gggaacatatgtagaaggcataaaagaagaagtgtttatctccagggcatgctcttttattgctgccggtgaagagcatcgccatgtt
 gttcaacaacttcaacttatttagcagccgaagtcacacatttgcattgatgattgaaagcagcgctcatggtgacgagtatgatggagtc
 atgtactcacaacttaatttgattgatctagctgggtcagagagttctaagacagagactacaggactgagaagaagggaaggatcatacatt
 aacaagagtcttttaactctgggaactgttattgggaactcagtgaaggagagcaacacatataccataccgtgattctaagttgactcgcc
 tgcgcagtcacatcaltgagtgccatggccatgttctactgattgcacaatcacccctgcatcgagtaacatggaggaacccataatacatt
 gaaattgcaagcagagcaaaacgtgttgaatttatgctgctgtaaccggatgatcgatgagaagctttgatcaagaatatcagagaga
 aatatctctctgaagcaagagcttgatcagctaaggagaggattgattggaggtgctatgcaggaagaatcatgattttacgccaacagtt
 ggaggaggggccaggtataaaatgcaatctcgctggaggaggaagaggaagcaaaagctgctctgatgagcagaatacaacgcttgacc
 aagtaatacttgtttacaaagaacaatactcctgacctgacagatacgtctatgcagcgccataattcggtaaatgaggaagataaagt
 aagcacatcacaagatagttctatgctgtccaaaatgatagtccacgaaagattcgctatcatctgcttcgccagatgctgtggatgagtc
 aatcaactaagatgtgctagtggagaccattcttcgattgctgttctggaccgatgaaatgcaggaggaggtaccgcacgcatcaaat
 ggatctactgattgagcaagtgaagatgctgtggtgaaattgttttggtaccagttcgctgaaaagactgattgagcaatctattgaagacc
 ctgaaggggacgaagaatcaaatagataatttagagcgtgaaatccgggagaagagaaggcatatcgagccctggagcaaaaacttatgg
 aaagcggtagggcatcagttgctaattgcttccatgatggatgcaacagacgattaccaagctgacagctcagtgagtgagaaggctttt
 gaattagagttaagatcagctgataatcgctccttcaggagcagctacaacagaagaatgtggaatcaatgattacaagaaaaagttcta
 cgcttgagcaacagctgacaacaatactgaggcatccccgaacaatgcacagaacatgagcttcatgattgaaatctaaactcagttg
 aaggaaagcagaaagtgaagagcttaatacagagcacatgaaatcactgaggaaaatcgtgaattgtcaaccagaacagcacattatgtg
 aggaagtgcctatgcaaaaagaactggcatcttctgctgtgttgaaactaagaatttggtgaaagattaccaagctatctgtacaaatgct
 aagcaggcaaaaagagttaataattgctcaggagttggcacactctagggttctggtagaaaaggccgctcagctggcaggggcagggac
 gaggttggaaacttgagctttagatgaggtatgaagatggagctacaggctagaaaacagagggaagctgccctagaagctgctttgg
 ctgagaaggagcatctgaagaggagtacaagaaaaagttgataggcgaagaaaaaggaactatcttggaacagatttagcaggaat
 gtgggtacttgttctaagctgaagagaggggctttaggcataatctgacttgaatgttgatgatcgtatataatctagctgacattactaacgg
 tacaaggaacaaagctgacaaaaatgtagctgtagttgagaacaacttcagataataaccgtgaagtcattgactgcagaggaataca
 gaaatccagaatttgaacctcttctgttcgtctcaaggctaaaattcaagagatgaaggagaaggaaactgattccttgggtgacaaagatg
 gaaattccatgtctgcaagatgtttgaatctgccactgctgcagtttgccttgcctggcattttgcttgcgaagccttgcctgcctgcct
 gtgcagaatgccctctatgccgtacaagaatgcagacaggataattaccttcacctaa
 >26317 OsAAK72891

atggaggccgtccagccggtgttgagtgacagattgaacccgttgatccaccacagatcagctttccagccagtggtgtgcagcctca
 agatcatggtgacagcgttccgggctgtgcagtgatctttattgacaccaggggagcgttagcagtgatcaatgacatctgaagattc
 accggcactgacgccgagatggtgtccatcaagtcacactccagctctgacaactgcttgaaggatcgaagagcggatcgtggtca
 gaccggcatgtttcaatccaaatgggcagggtcaactatgcagaattcatgattgatggaacaggagcttgataccaacttgataggctc
 aagggggatgtaactgggctggaaaacttgcgttgcgggataatgggtatataattggtacacacctgggcatgtctcttgatgtgatgctga
 ttgagattgatgagaggttcaatgccttgaagtactgtggccacagtgttgcgaaggctaggagatggacagctcgtcggtgtctgatct
 gcagtgaggagcatgagttgcaattggaggtcatcaacattacaataggagaattcatcagtggttgaagaggagatggagagggaagctg
 tatgagcaaatcttatgacaaactccatgagtaaaattggcaggatgccatagctcaatttgccagcatcgcgcatgatctgggtgctcttc
 taagctattgttgcctcactacaagaatcacatattctcacagcaagcatgaaactcaagcaacaggagcaatagggtgaagtacaacat
 ctltggaagaaaaacaaggagatcattcatctcgcggagggaataagagtttcagggaacagaagtcattgttgccttggaaaaat
 ctgactccgtcacctgaatggtatgagtaaaaggagatcataactattttaagtcgaaatgagcaaatgaagaggatgcatgaattgga
 ttgcaggagaagaccgaggagctcttcaattcaaacgggagaagggtgcttacttcaagaatgatgtgaatttgagccgttaagaa
 agaaaattccacagattattttagaatggacaaatcatttgaagaacataaagatgccttcacttgcagtaaatgatggactggatgaa
 agatgcatacagcaaaaagaatcattcttattacgaaatcagcatcttagaggttgccttgcagataatatgaaggatgtaaggaatt
 atcatctcagctatctgaagccagtaaaagaatgtctattcagttgtcatcagaagacgatctcctgagacaaattgccaaaattaaagaagaa
 tatgaggatctccaaatcgaagccggtgttagggatggaggtgatcaaaactattacaagaagttgcttgatgattctatgaacagcatgcatg
 atgctgcaacgaattttagtacagaactgtcttcttgaagctatgatatctgaaaaggagaaggcttgtgttgcataatgaagaaaccga
 atgttaaaagaaaaatagcagagttagagcaatgttgatccaagataaacaggaagatccagaagtcataagcaagagagtagacagaga

Figure 8
Page 3 of 14

tcattttcgtgatatagaggtggcgcctcacatatcaccaagaagatcacatgagactcccaagcaagatatgcaatatgatgaactgtcaa
acttaactcaagtttagaaattgcttcagctgcattgaaggaagttgaaaaataatcgattataatggtatttcaccaaaaaatgagcaaga
aaagcaactggagtgcatittaatatctattatgaagttgtcaaaggaatttgggagatcgagcaaaaattatcagtagaagaagtgcaagc
aggtcagaggatttgagtgatcattgcaaccacatggtcagacaagccgtgtactaacaataataggcctttggtacaacaaatgctagaa
acaagggcttctgagctccagaaggtgaagccaaggtgtcattttggcgacaaggttaattcactttaaaccttcttcagaagatatatgt
aactctgatcgctactcccctacattacagcaatatcttgggtgctagatgcttcttgaagacatgcaagcttgtgcagcctttgattaa
>26539 OsPN26539

atggctcgcggcgccgcgcatcggcctccaccagtggcgggcgccggggggagcgaggggtccccaccggccgcgcccgcgcccggg
catgcaggggcgccgagcgccgccacgcccgcgcctcggcctccgcgtccacgcccggccagcgagaccaccgtcgcccgccgcc
tcgacggcctcgacatccaggcgacgacgcgccctcgtcgcagcccgccacgagcaagaagaaaaaaggggcctggaacacgt
gcaacgggcccctgacaaggtggccgtggattgcgcaatttagtatgaaagtttgtgagaaagttgaaagcaaaaggagacaacacac
aacgaggtggcagatgagctttagctgagttgagaccccaacaataatttgcacacatgctgacacccctaacacaccacaattt
gatgagaaaaatatacgacgaaggtttatgatgcattgaatgtcctgatggctatggatattatataaggataaaaaggaaattcagtga
agggcttgcctcggacaagtatgagcgatgtgaagaattgaagacagagatcatcgactgaaaggtaggatcgacaagaaaaatgcat
attgcaggaggttagaagatcaattttaggtcttcaaaacttggcacagcgaacgagcgagctttatggttcaggaaatgtccttcaggagg
agtggcattgccatttatattggtgcagacacgtcctcatgctacagtagaagttggagatcagaagatatgcagctggtgcattttgattca
atagcactccatttgaactgcatgacgattcctttgtactgaaagcattgggggtctctggcaagaaccagatgatacgcaagcctgggttg
aaatggaggtgagtgctcaaccacacatctatcatcaatcacccaagttgcgaggccaacggagtlagactaccaacatcgcccccta
ttccgggtatacttaaagggcgtgtcaagcatgaacattag
>29946 OsPN29946

accatggggacggcctgcaaggagggtcacacattgggattatccgcgtgccatggccacgctgttcgacaagatcgacaagctcaag
aaccaagtagagttccagctgcgcgtttcgtttattgagattctgaaagaagaggtgagggatttgcctgatcctgctactgctgctgttggca
acttgagaatggaatggacatgcaaccaagttgtcgggtccaggtaaacccccctgttcagatccgggaggcgtcgaaagggtcataacg
ttagcaggatcgaccgaagtgcattgactacccagaaagaatgacggcatgccttgagcaaggtatctctgagtcgtgcaactggaagta
ccaacatgaacaaccaatcaagtcgttcccatgccatcttcacaatcacattggagcagatgcgcgaagcagatcccatcagcttagatg
gaatgcctattgaagagatgaatgaagactatctctgtgccaactccacttagatgctgctggtatgagaaaaaggagg
tgatggccttcggttaaggaaggtgttcataatcaacagaggacttctgtccttgcaatgtcatcagtgctcgggtgatgagaaaaaggagg
aaagaaggcgctcatgttccttaccgggacagcaaaactcactgccttctgcangactctccaggtggaacagcaagactgtaatgatag
cctgtattagtcagcagatatcaatgctgaagaacactgaacacttgaatatgtaaccgcgtcgttaacatccagaataagccaattgt
aaatagaaatcctgttgcgtgatgagatgaaaaggatgcgccagcaaatgaatacttgaagcagaactcgttcagctcgaggaggagttg
tcttagatgatgttcagggtctcagggaaggtatcaatgctgaacagaaaaatgaagaccttgcagggaactgtatgaccttcgcaacc
atggttacactgatccttgaacctgaactgcaaaaaattggaactggttacactaaagggtgaagggtcaaaaagaagcttgcaaaagtaca
gaaccatttgatgtcccatgactgactcagtaagagcagggcagccctaaagatattgatgaaagtgccaaagaatgggaacacacaa
tgctgcaggatagcatgggcaagagttgaatgaattaaacagacaactggagcaaaaaggagctgagatgaaaatgtatgagctgacac
tgttgacttaacaacactttggaagaaacttttgagcttgaagaagagaaaaggagctgtacagcagcaagaaagggacagattgttag
ctgaagttgaaagtctaaatgctgatggacaaacacacaagttgcgagatgccagctgcaaaaaattaaaaacccttgaaacacagcagatt
ctagacctcaagaaaaagcaggagaaccaagttcaacttctgaaggaaaaagcaaaagagtgatgaagctgctaagaagttgcaggaaga
aattcattctataaaggcacagaaggttgaactacaacataaaatcaacaagaagcggaacaattccggcaatggaagggtaccctgtg
aaaaggaaactcctgcagcagttgaggaaagagggggcgccgaaatgagtagaacgcccaaaactcaagcacttaatcagcggcagaag
ttggtttgcagaggaagactgaagaagctgcgatggctacaaaaggctgaaagagttactagaggctcggaatatcaggacgtgaca
actcag
>30852 OsPN30852

atggctcgcgcggcgccggcgatgggatggggaacccgggtgaacgtggggatcgcggtgcaggcgagctgggagaaccgcgagttcatc
tccaacatctccctcaacgtccgtcgcctctcgcacttctcctccgattcgaagctactacgaagagcaaaactgcgtccttgatgagaagc
tggacatcctagagcggaactggaggtgcttgaggttcaagtgcagcgcgaacaaccaatccatccgtttcaactag
>31182 OsPN31182

Figure 8
Page 4 of 14

atggtctccggcgtcgcaccgccggacgacgacggcgggcccgcgcgcctcgacgttcagcgcccgcgagccggccggcgc
gcggccgtccctggccacgccgccgccctcgggcggagcgcaatccgcttcgacgagcggcgggagcgcggtccccgtccagcc
gcagcgagcagcatgtccccgcagccgcagggatggcggcggggggcggcggcctctactccgattagtgaataccttcctccg
cctcaacgaccttgacatccacggcgacgatgcgccttcctcacaggctccaacgagcaagaagaagagaggagcacgagcagtt
ggcctgacaaagggtggcagggggctgcgccagtttagtatgaaggtttgtgagaaagttgaaagtaaagggagaacaacatacaacgag
gtggcagatgaacttgttgcgaatttcagatcccaataacagcatttgcaccagatccggataatcccaatgcacaacaatatgacgag
aaaaatatacggagaagggtttatgatgctctgaatgttctgatggctatggagattatatctaaagataaaaaggaaattcagtgaagggt
tgcctcgaaccagtataaatgatattgaagatttcagacggaactttaggactgaaaagtaggattgagaagaaaaatacatatttcagg
agctgcaagaccaattttaggtatgcaaaagttgatacaagaaatgaacagctatatggttcaggaaacattccctcgggtggagtgc
taccatttatccttgttcagacacggcctcatgcaactgtggaagttgaaatacagaagatgcaacttgacatttgcatttaatagcacacc
attgagttgcatgatgactcattgtactgaaagcaatgagttctgtggagaagaacaatcgacgggtattcatgatctaatttcaaatggagg
tgagagctcaagcatgccaaatatttataggcagcaagtgcaacacgtgcaagatcaactaatggtacagctagattgccaagctcacc
ctattccaggaatactgaaagggcgagtgaagcacgagcattag

Figure 8
Page 5 of 14

agactgagacaacaaagtttatcatgcaatatctacgtatgttggtggttagagctgctattgacgatcgaactgttgaaacagcaggttcttgaa
tctaaccttcttctggaggcatttggcaatgcaaaaacagtttagaaatgacaactcaagccggttggcaaaatttggagatgcaattgatg
caaatggttagaataatctggggctgctattaggacctatcttttgaaaggtcccgtgttgagcagattaatgatcctgaaagaaatttcatgctt
ttatcagttatgtcgtctggaaggatgcagaactgtacaagcttggcaccaggaggatttccattactgaaataaaagtaagacatatgaat
tggaagggaacaaataatgaagacgagtagtggaagcaaaaagggaatggatattgtcggaaatagcagaatgatcaggatgctatatt
cggatactggcagccattctcatcttggcaacattgagtttctcctggaaaggaaattgattctcaaaaattaaggatccaacatcaaatctt
acctcaaatggcagccaaactattcatgtgtgatccagatctgttggtctcaactctatgtacacgcgccattaacactctcgaaggagcaatt
ataaaagcattagactgttctgcagcagcagctaatactgtgatgctctagcaaaagactgtatatgaagactatttgactggcttgtgagaatatt
aacaatctatttggtaagatgttgattccaaagtacaaattggtatcttgataattatggtttgagagcttcaaaaataacagttttagcaatt
ctgcatcaacttcgcaatgaaaagcttcagcaacattcaatgagaaccgattggaattattgcactgctggatgaagcttgcattgttcca
aaatctactcatgagaccttgaacaaagatgttcaggaatttcttctgcacataggttgagaagacaaaattctcggagacagattttact
atctccattatgctgggaagggtgacgtatcaaacagaatcattttggagaagaatagagattatgttgcagagcactgcaatctttatctt
cgtcaagatgcccttagttctggacttttggatattaccagaagaatctttgagatcgtctacaagttctcatcagttgcttcaagatttaag
caacaactcaagccctcatggaaccctcaactgcagagagcctcactacgtacgtcgtgaagcctaactctgtaatcagcctcagatt
ttgaaaatcagagtgtcctgcacaaactacgtcggagggttttagaggctgttcgcacagcttctgctggctatctaccagaaggactta
tgctgaatttggaccggttgggtgttctggttcccgagttgatgcttggaaagttagatgagaggcattgacaaagggaaattctgaaaagat
gaagcttgataatttcaacttgtagcactaaggtttcttagagctggtcaaatcgcaattttagatatgcgtcgtgctgaagtttggagaac
gctgcacgccacattcaaggtcgttcagaacattcattacacgcaaaagatttgtgaaaacaaggagggttcaatttctatacagcatatt
gcagaggctgttggcaaggagaagatatggttaaaagagaacacgagggtgtataattgttcaaaaatgtcaggagatggcggtt
catcggaattaccagcaatctcattcagcagcttcttattcaatcctgtattcaggagttatcgtcgcattatttctcagttatcaaaagca
gaaagctgcttgggtgatacagcttctgtggagaaaacggaagggtatcattctttccaacaatacaggcagccactgtggcaattcagttg
gcttggagacagaaaagttgcaagaagagaactaaggagactcaaatggctgcaaatgaagcaggtgcactgcgtgagggcgaagaataa
acttgagaaaaagttggatgatcttacttgagactaactctggaaggagactgcgggctgctggtgaggaagcgaagtcagtagaaat
taaagcgcgacaaactgatagaatcattaagtgcgaatgtgctgctgcaagtcggctgctcaaaagcgaacacgacaaaaatctgctactc
cagaggcagttggatgattcattgagagagataactatgttcggagtagcaagattatgacagcagaagcagaaggagaaactccaatc
tgaagaacttagttgaatcattgtcaagaataattcatcacttgaatatgaactcacctcagctcgtaaaggtagtgatgtacgatgaagaaa
ttgaaagacgtagagggaaaatgaaccatctccagcaaaatttgacaaattgcaggaaaaacttacaaatgaaaaatgaaatcatgt
tcttaggcaaaaggcattaaatgtctccgtgaacaatatgccatgactacaaaggcttttctcagaaattcgtacaccgattgggttc
caaatggcgagcagaagcacggatgaaacaccaccaccagcaaaatctcgttcaactccacagagtttaactagatcaaggagaac
caggatgcctgttgaaggcaggaggaaaatcatgaaatcttattacgatgcataaaggaaaatctaggattcaaggatggttaaccagtc
ccgatgcattatctacagttgcccttaccactggcgtgctttgaaatctgagaggactgctattttgatcatgtcattgaagccataaataatgtt
ctcaagggggaagaggctgacggtagattaccttattggtgtccaatccttctcattgctatgccttctgcagaagaatttacggtcaaatgg
attatttctacaccatcccgcaggctggtggaacttagggattggtgacaagatagtcaaacactcagatctccttcaagcttatggga
cgcatgataatcttgacaagtggatgctcggatccagccatatttcaaacagcaactaactgcgtgttgaagatttttgggcaact
tagggataatctgaaaaaggaaatataccacttcttagtctgctcattcagggtccaaaatcatcacgtgcacagcctggaaaagcaacaa
atcacctgggattggtgctcaaccaccatcaactccattgggacaatattgtcaaatcttgacctgctatggacactttacatgaaattat
gtgccatcattctttatagctaaacttatcactcagctatttctttatcaatatagagcttttaacagccttcttctcggcgtgaatgctgtactttc
tccaatggggaatatgtcaaaaggcgtgtgctggtgagaaatggattactgatgccacggatgagttgcaggaaacatctatgcagatg
ctaaattatagacaaagctgtcggatttttggatcatatcaaaaaagggaagaagcttgaggagattaggaacgaactttgccgaac
ttgagtgtacgcaaatatagagatatgtcaatgtactgggacgacaaatacaataccaaggaatatcaaatgaggttgttctgcaatga
gggaggaagtaaacaaagatactcagaatctcgtatcaaatcttttattagatgacgacttatgtataccgtttctacagaggatttgc
tgccatccctgcaatagattatgtagatatagaactccagaatctcttcatcactatgcacagctcctactcaagcatcatgatccaca
gcctgtctaa

>23363 OsPN23363

caacggaggagaaaggagaaagagaggagaggaagagatggcgagccaggggagcgggaggggaggagggcggttccctccgg
ggctgccgtcaccatcaccaccgcgccatgacggagacggaggacacatggccgtcggcaggaggaggaggtggcgcgccgt

Figure 8
Page 8 of 14

cgggcggagacggaggagcacgtgcagcgcacccctcctgccatcgacgccttactcgcaggtgtcggagatgctggaggcggggcg
 tgcgctgttcaagaacctcgcgcgaccttcgaggaccgctctgctcgcacacaaggagaggggtggagaggtggaggaggagatcc
 gggagctgcgcgcccgcgacgcccgaacgagcaggcacgctccctcctccacaacgccagctccacctcctgccaccgtccgcca
 cgaccacacctga

>23390 OsPN23390

accccagcacttctctatgtccatacctccggaacttctgaggcaattccacttattgatagggtccaggtagagggctttctgaaggcaat
 gcaaaaacagatccattcagctggaaaacgtggttttttgaagaaatctgtgggaccacatgttcgtgaaaagttcaccttggagatatgt
 tgtgttccaaaaggatccatttccaacgtcattgtctaaaaataagtagtgaccttgaagccgttcaattaagtattccatgtcactgaagta
 catgggcatcgattccccagcaataaagtttgatgaaagaattgaactgttgaagctttacaagcatacttgaacgttctgaactga
 gagatgagctcttgcacagatttcaaaacaaacacgtacaatcccgcagggcttggttaataagagcttggagcttatgtattatgtgc
 gtcatcatgccaccaagcaaaagatatggagcatacttctgtgagtagttcactacattgtctatggcgctacaactgattctgatgttcgagt
 tctagcacttaacacactaaatgcactgaaacgctcagttaaggcagggccctagggttacaattcctgcacgggaggagatagaggctctttt
 atccagccgtaagcttacaacaattgttttttcttgatgaaactttgaagagatcacatatgacatggcaacaactgttctgatgtgttgag
 gaacttctggcattatcaaaacttcagtatattctagttcagccttttgaatgccgaaaggttgaatggatccaagcttcagatgttgcaa
 tgaggagtagattggttagatgataacaaatattggtgacctctatctgaattcaaggcagccaaggatcggaacaaaggagaaatttta
 cactgcaaaactgtatttaaaaaacgcctcttccgtgagtcggatgaggctataactgatccaatgtttgttcagctatcttatgtccagttacaac
 atgattatattttaggaaactatccagtaggtaggatgatgtgcacagctatctgctctgaaatattggtcgaattggttctgtgacaatcc
 tgagcttctgttgatgatatctcttttagagagatttctaccagacaagttgcaattacaagagcaaaagcgagactgggagcttgacattg
 tctcacgctatcagttaatggaacatctgtcaaaagatgatgcaaggcaacagtttctcagaatcttaaggacttctccatagggaactcagttt
 tcttcagtgccggaagatgatgatccaataggacttttacctggaagaatcattctgggaataaacaacgaggggttcatttttccgtccg
 gttcccaaggataatcttcatctgcagaactgagagatatcatgcaatttgggagcagcaataaccgctgttttctttaaagtagagtcgctgg
 tgttctcatalcttcatgttgaaactaagcagggcgaggaaatatgttagcacttcagacgcataatgatgtgatgtccgcagatattc
 aaaagcagctctgctactagcgcggttttctcaaacgatgtttctcaaacatataaaccaccgaatattgaaatatagagaaactgtccaa
 gaactgtcaaaagcagttgaggaatctgaaaggaaagcagatctgtgaatgaggagttacagaagaagcaaaacaaagagagatatg
 caaaaagaattagaaggtctgagggatcccttgcaatctgaacggcaaaagcattaaagaagtaacaaatgatcttgataaactaaaatccctt
 tgtgatgaaaaggactccttctgagggcttactgatggagaaaactagactggagaccagattaaaaagtggtcagggccaagaaagca
 gtaatagaacaggagtagatcaggaaatcatttgagagagatacttcccaactgtaggcactgtcaataatagcattgagatgttagcaaaact
 tgaggaggaggttaaaactctgtaagaaggagcttgatgcacccaagggaattatcaaaagaaatgaacatggaataatctgcttgaccagaa
 ggttcaaaaggcttgaagagcaaaaagtgaagagaaaagtaatatggaagagatttatgaggatgaatgttgaagctgaaatctcgtattg
 ctgaattggagcaaaaactggaaagcagaacacgttccctgaatgttactgaatcagcttgccttaagaaatgtgaggtggatacattgc
 aaaacagctcacaagaacttgacgagttacgagagttcaaaagcggatgtggacagaaaaaacacagactgctgagattcttaaaaggca
 aggagcacaattgattgagcttgaatactttataagcaagagcaggttctgagaaagcgttattataacacaattgaagatatgaagggaaa
 aataagagttttgtcgtcgtcctctaaatgataaggagctcattgaaaaggacaagaatattgttgcagccctgatgagtttacagttgc
 acatccatggaaagatgacaagtcaaaacaacatatatgaccgtgtttttagtctaactaccaagaagaagtattgaggacacgaa
 gtatctagttcaatcgccgttgatggatataatgtttgtatattgcttatgggcaaaactgggtctggaaaaacttttacaatttatggatctgaaaa
 caatcctggtcttaccacaagggtacacctgaacttttaggggtgataaagcgtgatgtgcacaaatattcatttcttgaaggcatatatggtt
 gagctttatcaggataatctgtggacctgtgttgccaaaaacgcaacacacacaaaaattggaaataaaaaaggattccaagggggtgtt
 actgttgaatgtgacagttgtgaacattcaagtttgaagaactgagggctataatttgaagaggttccgagagaagacacaccgctggaa
 caaatatgaatgtgagagctcaaggctctatttatttcaatcattattgaaagtaccaacctccagactcaatcttatgcaagaggaaagc
 taagtttctggacctgtggttctgagagggtgaaaaagtcgggctcagcaggaaacaactgaaagaagctcaagcatcaataagctct
 ctttctgactggctgatgtgattggggctttatctctgatggacaacatataccttatcggaaccataagttgacaatgctgatgagtgattctct
 tggaggcaatgcgaaaaccttgatgtttgtgaatgtctcaccagcagagttccaacctggaggagacttacaattcactcatgtatgcttcacga
 gtgcgttgactgcaatgatacagcaagcatgttccccaaaggaaatcatgaggttgaagaagtaattgcttattggaaggagcaagct
 ggcaagcgtagtgaagacgacacctggaggaaatacagggaagagaggacaccaaaaggagaaagcagataatcgttactagctga
 >23416 OsATPF

Figure 8
 Page 9 of 14

atgaaaaatgaacccattcttctgttttttagctcactggccatccgctgggagttcgggcttaataccgatatatttagcaacaaatctaataaa
tctaactgtatggttggtgtattgatttatttggaaagggagtgtaaaagatttagataatcgaacagaggatcttgagtactattcga
aatcggaagaattgcgtagaggaaccattgagcagctcgaaaaagctcgaattcgattacagaagtcgaactagaagcggatgagtatc
gaatgaatggatactctgagatagaacgagaaaaagcaaatltgattaatgccacttctattagtttggaacaattagaaaagtctaaaaacga
aaccccttattttgaaaaacaaagggcgatgaatcaggtccgacaacgggttttccaacaagccgtacaaggagctctaggaaactctgaata
gttggtgaataccgagttacatttccgtacgattcgtgctaatttagcatttctggggccatggaatga

>23484 OsPN23484

atggatccaatcgcgacccatcgccgtcgtcccgggcggtcggtgaagcggcgccctcccgcgatcgccggagctgtcgccgaagg
cgtggggaggggaggcgccggagctgatccggcggtggaggagctggaggagcgcgcgcggtgctggggcgaaaaaggagg
cccgaggaggagcgcgcgggagctgcaggccgagttggatcgggagcggcgctcgcgcggaacggcgaccagcgaggccatgct
catgatcgagcggctgcagcgggagaagggcgccgcgagatggaggcccgcagttccggcggtacgccgagggcgcgaggatc
gcgagcgtgaggtacaggaagagctgcctcgtctccgacctcgccctcctaccactcccgcctcagtcacgggagtcgacctg
attccttctccgacgagggcgaggaggagcagcacgacgaggaagatggcgaagaggtagaacagatcgacagggcgcgctacaga
cagatggaagcagcggcgcgactcaatcggtggcatgaggtgaaagccatggtcgacgatgatgaggaagagcaattcaccctgtg
gagaaggagttcgagtacacgggtgatgtgaggtcgcgagctcgacgacgaaggtatccggggcggtggtgttggggagttcatggg
ggaggggcaatgcagctcgagggggattgtacgccagggtggaggctctggaggcgacagggcggaatcgaggaggaaatcgag
cattcgagcagagagagctcagctggtgatggcgaggcgatggcgcgcggtgtgccgggaggtggtgtgaacagaaagcca
agaaagtggctgcctcaccacgaagcttctgggtgctagggggttgcaagtgggtgctctcgataatctttggagaacagaagtccacag
ccagggtatccttcggtctgtaactacgttcttggcttctactgctgctcgacagatccaccatgttgagtcctggcgctgctgcctagc
ccacggcaatga

>25358 OsAAK39589

atgaaagagcccgcgaggtgaggacggcgcgggggcgcgcgggatgggggtggagggtggaggaggaggaggagcc
gccgcgatcgcgacggtgaagcaggaggagggcgaacgcggtgctcgggcgagggggtcccgccttgcctatcggggagctca
aggaggaccacgaggtcgcgccggggagcggcggtgaaggcgcgctggtggggagaggaacgggagtcggatcgccggatgccaggg
ttcgtctacagccaagagagtagtcagcagtttcatcccatcatgatgtgcaatggacttaataaagtgtcactggagttgacgaggaa
ggccgttctcgccagaggattcttcttctgtgccccaaagatacataagcgcaatagaagaaatcatgatgacctgatgcataattataattg
ggcccttgtctccaggagagtgagacaatgtagatccaaattctagttcgtccaaagatgcattgcttgaggaggcttgcataaagtatgct
gaagctaccgcttctgccaactctttagatgcataattataactgggctattgctattgctgatcgggctaaatgcgaggtcgactaaagaa
gctgaagagctctggaagcaggcaatactgaattatgagaaggctgtccagctaaattggaatagtcggcaggtccttaataattggggtctt
ggactacaggaactgagtgcaattgttccggctcgtgagaagcaaacatcataaaacagctataagtaagtttcgagctgcaattcaactg
caatttgattccatcgggcaataataatcttggacagcttctgtatggttagctgaggatacaatgaggtctgggaagccagggtgtgccg
ctagcgagtttacagtcagctgctatctatgttgacgtgcccatgactgaagccaaattactcggtctaccgcagtgcttgcgattagtc
gttcaatgctgccttgcctatctcaagtgggataattgattgctctccagaaatagtgcattgcaccacacaaagaatgggagagggtc
acagtttgtttgaaccatgaagaactccagcaggtcaatgcctctgaccagcctccatcacatcacctgggcatgtggacagcggcagaa
agcttttcaggatagttgttcagacattgttctgttcagcatgcgctgatctaacctgccaccgggtgctggccttcgcatcgatactattc
atggccctagattcttgttgctgataactgggagaccattgacagttgggtagacgctatagcctagtctacaccatatttgcaagagggag
gagtgatgctctgtgggattattactggtga

>25381 OsAAK20062

atgggtgatcaacctagaattcatcagggcgatcggtacggccgatgagatccttctactggatcctctcacgatcgatgtgatccctttgtga
gcaattgacacatcatctccctctcaagaacctgggtgtgtgggaatggtcaacctgggtgtgatgaccatggggaaaagcatgatgactcgc
atggagatcaggtgccccgccttaacgagggccaccggagcagagcagagtgccattgaattccagggtgctggagctcgactcgaga
ccgtgtgtcatcattcgacgttaacgtgtctgtgttgaaggcgcgccactccggtacttgagggaactgaccaagaatgcagcacaagg
aatctcgatcgtgtgcgaactctcaagagtgatcttaccggttgccttgcctatgtgcagaaggtcagagatgaatagaacatcttctggatg
ataatgaagacatggcacatctgtatctaacaagggaagcaattacaaaaccagcaagttgaggctttaatcatcggtgcttccaatagcat
tgttctggaggaaacagctgtccagggtgaacaatagtttccggcgagcgtgagcatcgctacaagcatgcatttgataatgacgtgga
agacctagaatgttgcttgaggcttacttcatgaattggatggaattcgcaacagaattttgctgggtcgagagtatattgatgatacagaag

Figure 8
Page 10 of 14

actatgtcaatatcaacttgacaaccagcgtaatgaacttattcagcttcagcttacactgactattgcatccttcggcatagctgtcaacacctt
catagctggggccttttgcgatgaacatccagagcaaattgtacagcattgatgatggtagcttcttttggccatttgttggggggcacctcatcgg
gctgcttcatgatctgcatcgtcttattatggtatgccgggtggaagaagtgtctcggctcttga

>26210 OsAAK38489

atgctgcgccgtcgcctgcctccagcttctggaagactttgatctggagaaggaaagaggttggttggatgagcagggtttaagatt
gcagagaatcaagagaccagccagaaaaacagacgccggttgcctgagagcactagggtattcaagaaggcgtcttcagatgataagctt
agttattcaattcttctgctgaagagtattcaagagggaagttgataatctcacaagagagcaaaattggagaaaatgcatttctcaacatctac
cagaagctctatgaagcccctgatccatcccgctcttctgctctatggctgagcaagatcagaaaactatcgaactggagactgaaaatcgg
aagatgaaacttgagcttgaagagtaccggggcagaagccgccacctaagaaccaacaagcaacaattaggcgacttgaggagcggaaa
ccgccaaactagaacaacagatggaggaaaaagtcagagagatggttgaaatgaagcagaggaggttggcagaagatagccaagaaaactc
ttgaagccctgaaagacaggggagcgggcattgcaggaccagctaagacaagccacagaaagtgtaaagaatatgcagaagctacatgaa
tcagcacaaagtcaattgttgagcttctgactcaatcagaggaggaccgggctgcaaagg
aaactgaagtcaatcttctgttgatgaagttgaacgggcccgaagcacgttagttagtcttgagagagaaaagggtgattacgatctcagtt
gcaaacaacaaatgaagatgcaaccaatagtagtgactacgtggactcaagtgtatatacttgagaggtccctgaatgccaaaggagaagatca
tctcagagctgaatgcagaactgcgtagtattgaaaatactttatcttctgagagagaaacacatgttaatgaactgaaaaagttgactgcttgg
ctcagtgagaaggaaaatgctctgacagagttgaagaaagaactcaagaaagaccaacaagaagactagtagatgatctcaagaaaaag
gttcagatttgcaggctgtagggtacaactccattgaagctgaagattgggagctagcaacaaatgggtgaagaaatgagcaagttggaagc
ttgcttcttgataagaaccgtaaaaatggagcatgaaactaacacagctgaagggtcaaaatatcagagaagtc aaatcttcttgaggaggctga
aaaaagattgctgagctaactgcaaaaggctgaagagcagcagaagttgatttgaaaactcgaggatgacatactaaagggctatagttaa
ctgataggaggacttcacttctgaatgattgggatcttcaagaaattggctcaaatgaagtagcagagggtactgatccaaggcatgccaccac
aagaccaagaccaaaagttccatgcttaaggtcatttgcaatcagaggggatcgtttccgtacacgtctacgtgaaactgaggaggaactaaga
agactaaaaagagaagtatgaaatgctagtgtagaattgaaaaaaactaaagcagataatgttcaactgtatgggaagattcgttacgtgcag
gactacagccatgagaagattgttccagaggaccaaaagagtagtcagaagatgttgaaagtgttcttcagatgttgagacgaagtacaa
gaaaatgtatgaggatgacataaatccttttctgctttttcaagaaggaaaaggatcaacgggtacaaggaaacttggttaagagacaaaatc
actcttagcagtgagcgttttctcttggtaacaaatatgccggacatttatattcttctacactattggattacatctccttgtattacattgctgt
acagaatgtcagcttggagctatctcagacgtctgaatacgttttctgtagataaaaaatttccagatatggagacgggggtggatgggaaatga
caggatgaggcgtggctgtcgcttggagccgttattatgtaaactgttgatcactaa

>26688 OsPN26688

atggcttctcttctctctcccttggcccttggcgccatcttccaaagcgggtgccccctctcccgctcgcgcccgccgtccgccgcgcgcc
caccaggcgccgcgcccgtcggccaagatctcttgcacggatgggacccccgagggcgctctggccccgccgcagggcgccacat
cgtgcgcctcgagttccgccgccgctcgagagggactccgacgcccgagggccttcgagcgccaggtccgcgaggagcacgagcg
acgccgccaggagcgcgagggcggggtcatccctgacacggacgccggcctcgtcgagtcttctctcgacacggagggcgcgcgagatc
gaggtagagatcggcagggtccgccccagggtcaaccagcccttctcgactacatccagcgcgagatcgccagatcaagttctcaatca
cccgaacagcggaaaatggaggaccgattgatcgagttggaagcgatgcaaaaggttctgcttgagggagtggaggcctatgacaagttgc
aaaatgaccttgtcagcgcgaaagaacgcctcacaaagatcctgcaatcaagcgacaaaaaatcgacacttctgaaatggttgaaacggaa
tgagctcaacatgtccatactaacactcctcgatgaaaatatagcaagcgccaagaccaacaatcaggaagaagctgtggctttcatggaga
atgttcgctcatccatcctgaaatatattacagtatga

>29882 OsPN29882

cacgcgtccgcaaatgagcacattcttgctatggaaaaggaggtgaaaaattgcaggctcagctgaagcaagaatcattgctaaggcagc
aggagcaacagaaactttctgaagagtcctattaaggcaacaagagcaacaaaagctaactggagaacagctctatgctgcttcttggtt
gcggaaaagaaagatttggagaaaaaattgctgccttgacgaagaagcatcagatgaagcttctgaatttgctgcacgcaaggcatttc
aatggaagatagggaaaaacttgaaagccaattgcatgatfggcttggatgggtgagaggctagagggcagccgcaaaagtgttaatgg
agattgattcctaatacagaaatagaaaagctgtttgaggagaactcagccttgctacttcataccaagaagctgtagctgttacaatgcaa
tgggaaaaccaggtaaaaagattgtctcaagcaaaatgaagagctacgctctcacttagagaagttaagacttgagcaagctactctattgaaa
acaagtaataccactatccagccagatgggcagaaacgagaccagcatttcattcccgccagagtgtgaactgagaatctctccctaaggat
cagctcatcaaaagaacagagcagatccgaggggtgtcggcagaaataatgaaacttcagctgaactcagaaaagcggtgcaagcgca

Figure 8
Page 11 of 14

gaataacctcgcacgcctatataagacctgtgtaagaggcatagagagcaacttaataaaaatgaaacaagaacataatgcaacgatccagt
aa

>29942 OsPN29942

gagcatgaatcagatgaagagaagaatgaagttgatactagcaatgaaatcaaagagattgtcattgatgacatggagccaaaagtgttcca
ggctgggtttcttttatgtacagggacaatcttgttggatgatgatgagttgtctgcatcaagctcggactgctctatctttgatactttagctggga
aattattggctgctgctgacagatatgagttgccaagactacgggtattgtgtaactttactgtgcaagcatatttctgtaaacctgtggcaac
taccctagcattggctgaccgacatcatgctatggagcttaaatctgtttgcctaaaatttgcgtgtaaaatctttcagctgtaatccggaccga
cgggtttgattacctcaaagacaactgccccgcgtgcaatcggagatactgaggacggctgccgggtgtgaggaggaatgcagcagcg
gagggaagagccagagcgtgtgggggcagctctcggalggcggcgataccagcgggcgcagggtgaggccaagggtctga

>29956 OsPN29956

tttgagattgaatgggagctgattgatgaaaagaagaggagctacaaaaggaagcgtacagaattgtgagaacgaagagcaataact
gagtatctgaagaatgaatctgatcatcaaacaggagaaggataatctccgtgttcagttcaaaagtaattcagaacacactctctctgaac
acaaagagttcatgagtaagatgcagcaagaacatgcaagttggctgagtaagattcaacaagaaaggcaagatctgaagagagacattg
atatccagagggtggaattgctaaattctgtaaggcaaggcagatggaaatagattcttattgagggaaggagggaagagttgagcag
aaaaaggccaagggaactcgaacacatcaattctcagaaggagatgatcaacacaaaattagaacatgttcagttgaattgcagaacttaa
ggatgagagaaaaagaagctactttggaacgtgagaggagagagcaagagttgtctgagataaaaggcactattgaagccttgaataatcaa
cggggaagaagctgcaagagcaaaagaaaactattacattcagaccgagaagcaattacagtgcaaaatcaacaacttaattgttggagaact
gaaaattgattctgaaaataagcaactgtctttgttacaacatgataagtcaaagcttggagtgatataaatgtgaagacaatcatcatgata
actccattctcaccaaaagcaacgttttgaaggaaactagacctttccagctcacaaccgatttcttgggttcgaaaatgtgctcaggtg
atattcaaacggctcaccagagaagagcgttagccatgatcaattgtccagaatggtgtgccaaagaaagttggagactctgtgatgttga
gatgtgaattggactttgcaaaagttggtcaaaagaggcttaatacttggttctgtgaccaaaactgaagttttagccaaaacgaaagca
cagaaggagtactattcagaaggttaattggaggggaaatcacttcaactgcctgtcagctctagaagagaatgtccaagaatgaacatg
atgaagcccc

>29957 OsPN29957

atggcgaagtcgagcgcggacgacgccgagctgcccgcgcgtgcccgcaggccgtggcggcttcggggggcgcgcggcgaggagg
tgtcttctccatccgcgtcgccaaggggaggggcatcttcgagaagctcggccgcctcgccaagccccgcgtcctcgcgtcaccgttaa
acaatcgacaaaagggtgaggcagccaaagcttttctccgagttattgaagtattcatccggcgagttacttgagccagccaagctttacaagtt
gaaacatctcacaagggtgaggttatttcgaatgatcctagtgttgacgtttgtattgggtttgataaccttagaagtcagagttgtcccc
tccacaatggacaatgcgcaacattgatgacagaaaccgctacttttctatcctgacctgtgcaagagatacttagctatcttccaaaag
tcgttgaattgattttgtggagctagcccttgggccaaggaaaatacagtgactctggacaaccaatcgagcaccgaagatggacaaga
aaaatctgtcacaactcaaatgagaggaaagtaactgtaactgttgaatgatcttgggtcccaagcaaaagatgaagaggaagacatg
gaggcactcttgacacatatgtcatgggcataggtgaagcagatgcttctctgagagattgaaacaggaactgttgccttggagcagcc
aatgtctatcaactactgcagagtgaaactttaatagatgaggtttgcagggtctggatgctgctagtgaaccgtagatgatatggatgag
ggttacgatttttaatatgaagctcaggcatatgagagaagatattgcacgattgaatcacgcaacaatggcttggagatgcagtcctgtaa
taacaaaggactcgttgaagaactagaaaagttgcttgatcgcttgcgaattccacaggag

>29958 OsPN29958

tcattgcagaatgaagtcagtgccctagagaaacaaaccttatcccttgccaatgattgttgcaatcaataagctcaggatggaggaaaat
gcattatcaactcaggttttgaagaccaatatgagatcgagtggtgatcagaatacagtacggacagtaaaagacatggagctgcagaaatt
gcatgggaccatcaaagcactccagaagggtgttacggatacagccgtccttctgatcaagagagggttatttcaatgccaatctgcaag
aagcgaggaagcagattgaggtgctgaagctgaaggagattttgatgatgacttgattgaaatgaattatgagcaaatgctgaaagacatt
cagcttgacctcatccaaattcttctggtataaaaaccggttcccttggtcaggccaataaaaactgtagcacaggcaaatgagaagatgcttg
actctcatggcatcgttggagctagcagtagccatgtacgcaatgattgagaccgccgcaaaagttagtctgttggagaggacaattacaaa
aggcctccctctgaactgatggtcgtgaaagaactcagcattgacaagcaagagttaccaaggtcaatcaccacggagccgcaccaagag
tggaagaacaaggttaatcgaaagattggcttctgatgcacaaaggctcaatgccctccaatccagtattcaggagctcaaaacaaacacag
aggcatcagaagggtcgcgagctgagagcgttaggtaccaataagagaagctgagggtctcatcactcagctgatcagatgaacggta

Figure 8
Page 12 of 14

aactgtccaagaaggctgaggagttcacatctgaagatggcttgatggggacaacattgacttgaggagcagacaccagcgcaagatcat
g

>29961 OsPN29961

gatagtgcctatggaccacaagtatgggcaaaaagatgggtgctctgatgagggttctgtaggtgttcctggtaggaaaaccaaagagacagt
aactgctgctgatgcaattattgatgcacttgacactgccgaggaagaagtaaacgccttgatcagcaccaggaaggacaaaataatggg
aatggaactacattcaacccaatgttataatgcaagggcagtcacatcggattatgtttgaatgtatgtttcaaatgttcgtccaaatgacttgg
aacaggctctccttcatgccattctcagatgcacttaagatcatgtcttacttaaggagtggtctatggttcctttgaagggtgagctgtttgta
gggtttgcctagtgtgcttcaaacacaccacagccaattaactacgacacatctgcaagatcgatacttacagaactgaaaggcattcttta
cagtagagtaaaaggaatgcaaggatgctataggtttcaaccttgcgcaatggatcatataaaagaattgttgccatgagatcagatgctcc
cttcgagatgcaagagcgaagctgatggagatcaggcaggaacaatcaagacgatcgataggtcagatggggctgaaaaaaggaaa
aagaaaaagcgaagaccatccggtgagagctga

>29965 OsPN29965

ctactagaggaaaagcacgtacgcgaaaggcttggcttcggctgcgggaggtgaattgaaagcgttgcggaagaagtcaccaagctcatg
aaccaaaacgagaagccttgccttgagttggcatcgggtgagaagtcacccccgcgtagagccaacagtggactgagaggtactcggag
ggatagcatcagtagacgacatgagccagctccaagaagagacaacaacgcaggctacgagagagaaaaggcattagaagctgtgctta
tggaagaaggaaacaaaagggaagcagaactccaaggagaattgaggagtcgaagcaaaagggaagccttcttgaaagtgaacttgcaaac
atgtgggttctagtggcaaaactgaaaaagctcaaggccatgatcttgaggatgtttgacacaaaatatattggctcgtaa

>29966 OsPN29966

gttactgactctgttatgaagggaactataagtgtgtagagcaagaatcagctcgcagattgcatctaaggatgcagagatagcattcttga
acgaaaagctgcatcaattcagaaatagtgccctgagtttatctgagggccgagataaactctatgaagaaattataaccttcgccaacagct
tgtactcttctaagtcactcttgaattctgagtggggctttagatcacactataacaactttgaaaggcgcagaagatgaaagtaagcata
gagggaaatgaaaaatccttaaaagatggtataactaaggaaaatggttctaaaggctccaatgaagatattttattgatccaacagttctgaag
cacatggataggatgagttgtagccatttcaataaaatgatgaatcaaatgaacgcgagcatgattcgactttgcaagagaagacgga
agagatatttagactaaagcgggagaacctaaaaaagaaggacctaacccttgccatttacgcaataataaagaatttgagctcatgagga
agaaaatttggaaggtattacaaaattggatgaggtccttgggagaataaaagaacatccgcatcaagagtgatgtatttctggtcaaca
agac

>29967 OsPN29967

tctctctcctctcctcctcccttccaccccacgcgccccgcacccacccacgcgcccgcgccttaaccctaaccctaaccctaaccctact
cctctcgtctgcatatgctcagccccgcgtcgcacccaccggccatgcccaggagcgcgcgcgcgcgcagccgcagccggagccggc
ggttcacgcggggggaggaggccgccccgtcttgagaggaaggaggaggtgctgcccgttgaggagagaagatctcgaattagatgaa
tcacagtcacagcttatggggagacttcggggactcaaggaggatttgcgaattggaggagcagtttagatacacaagtgacaaaatacaa
aagcgaactctctgatatcaaaactgcgctgaatagcgaattgaacagctgagatcagatttcaagaattgaggaccaccttaagaagc
agcaggaagatgtatcaaatagtttgaagaatttggggctacaagatgctactgacaatgatgggaacaaagggaagtg

>29968 OsPN29968

atggagaagtcgccactggcgtggttcaggtgctggtcaacaatgaagacgtcgttgccataaagcagatgcagcatctcatactgggac
gattgcaagatagtaatgcagttctcacacatttgaatgaatactctgaacaatgcttgcggaggtgtctaatgactttgctgtaaaactcgtct
tctcaagtccatgaaggctgatcttgaccatatttctgaagttgagagcagatgaaatcagggtagcggcaacatatccagatgcttttccta
ctggcgcaatggcagagacgatggaccaaagaccggaccttgaagtcctcttgattaa

>29969 OsPN29969

atggagaagtcgccgcccggagacagccgcggcgccggcggaggtggaggcgcggttcaggtcgtggtcgacaccggggacatcgg
cgccatcaggcagacgcagcacctcatactgggacggttgcaagatagcaatgcagttctcacacattttaaagagtactctgaacaatgctt
tgccgaggtgtctaatgactttgtagtaaaactcgtcttctcaaatccatgaaggacgatcttgaccatatttcttgaagctgagaagcatgaa
atcgaggctagcggcaacatatccagatgctttcccgatggtgcaatggcaaaaacgatggaccaaagaccggaccttgaaggtccgctt
gattaa

>29970 OsPN29970

Figure 8
Page 13 of 14

tcgtcctccgaccaggctgcagaatttactgatatggaaggcgaatcatcagcagttacttcaccatttctgctctgacttcaaccacaccaa
atgaattggagatgactaacaagaattccaatgtggtaggcggaatgactcacagcaacagtatgccacattgactgctgcaaaagatgga
aataccaaaagttttacccttcgaattccgagcacttgaagtgtgccttgagtcagcctgtagaagcctagaagaggaaacgtcactctggag
caagaggcgtatccggcacttgatgagctgacttcaaagataagcacactaaatcttgagcaggttaggcaaatgaagagccgctggtgc
aatctctggtcgtgtgcagaaggtgagggatgaacttgagcattgttgatgatgaaatggacatggctgaaatgtatttgacagaaaagctt
actcgacaaga

>30848 OsPN30848

ccacgcgtccggccttggaaactttgagctttcaagatgccaaggaggacagataatgctgcttctgccaaattcagttgaaccagaaaaatca
gaagaatgtttggagtttgatgatgatgaggaggaggaggtagaagaggaggaaattgaatatgaagaaattgaggaggagatagaagag
gaggaggtagaggaggatgaagatgtcgtggaggaaagttgaggaggtatagaggaggaagatgaggaggaagaagaggaaatctgat
gaaactgaaggtgaagtaaaactaaaggtgtcaccagaaggatgttactgaaaagggaagcatgctgagcttctgctcttctcccat
ggttctgaagtatatgttggaggcatctccagtgtatcttctgaagatctgaagagactatgtgaaccagttggagaagttgtggaagtga
gaatgatgagaggaaaggacgatagcaggggatgcttttgaattttagaacaaaaggttggcattaaaggctgtcaagaattgaaca
atgcaaaactgaagggaagcggaataagggtttcttctcgcaggctaagaacaagcttttattgaaatgtaccccatagttggacagatg
atgatttcagaaaggttggaggaaagttgtccaggagtattaaaagctgatctcatgaaggtcctcaaatgcaaatcggggttatg
gtttgtgaatactacaacctgcatgtgcagagtatgaaggcaggagatgcttcccaactcaactagattcaaatgctcctacagtc
agctgggcagatcctaagaataacgactcggccttacttctcaggttaaatctgtatatgtgaaaatctgccaaagaatgttactcaagcac
agctgaaaaggctatttgagccatggtgaaattgaaaagggtgttcttctcctcaagaggaggtcatgataatggtatggtttgtcact
ttaaggacagatccatggccatgagggtctgcagaacacagagagatatgagcttgatggtcaggtcctggattgctcacttgcgaacct
cctgctgctgataagaaggatgataagtagtaccactacctagttcaaatggagctccattgctcccaggtatcctccacttggatatggtatcat
gtcagtaccaggtgcctatggtgctgctcgtctagtagtgcacagcctatgctgtatgctccaagagctcctccaggtgcagcaatggtcc
aatgatgttaccggatggtcgtcgtatgtgtacaacagcctggtggacagttgccgctggcttcgccgccgcagcaagctggac
atcgtagcggcagtgaggacgtcatggcggcag

>30854 OsPN30854

atgcagggggagggtgcatcagccgatgcagatggtgctgcgggtgaagcacccgtcgtcgtggcgccggcgccggcgccggcgccggga
ggaggaggccggcgaggcgtcgtcaggtcggcgctgctggtgttcaaggccaaggaggagcagatcgagaggaagaagatggagg
tgaggggagaaggtgttcgcagctcggccgctcagggaggagtcgaagcgcttgccttcatccgacaggaactggaagggtggtc
ggatccaaccaggaaaggaggtggaggtgatcaggaaggagatcacgtggtgaaccgtcagctcaagcctctcggcaaaacctgcgtc
aagaaggagaaggaatacaaaagaatcctggaggcgtacaacgagaagaacaaggagaaggcactgcttgaacaggctaattgagct
ggtgagcgaagtgagaggatgcgtatgaagaagctcagggagctgaacaagacgggtcattactctactag

>30899 OsPN30899

gttcttctgattctttgaaacggaagacatatgatgatgagctaaggaggaggagcttttgaactacttcagacggtttcagagtgttctca
aaagaaaggagggaagtggatattttcgacaagggttagcccttctgaaggtgttgatgaaggaccttatggtttatcaagaagaatagcctgc
aagaatgtggtgacttccacctgtggatttatacaggaagagctaaatcgcaagctagatggtgtcaggactgcaatgatttcaccaagct
aaagatggggatggatgggtgagcagcttttcaaccagttctattgggtgctgcacaagcctgaattgcctcatgcatatgtttgtgctgaa
agcatattttcgtatgcaccgagtggttcacctgtcagggaatgagatgccccgcgaacactcacaagccgagcttccatgtcaatgccag
cttgttaaagcagaatagtggaaggaggacacctggcgagagggtggaggagatccctaattggtgtaaacatggatggcggaatcg
atgaaggaggagttcttcagtggtgcagaatgcactgcagctcggcatgtttgaatcgtttggcgccgcaaaacgagcccccttccctggaa
gtggaagcaatgccaagggtagtaatagcagtagt

Figure 8
Page 14 of 14

aatatctggggctgctattaggacctatcttttgaaagggtcccgtgtgtgcagattaatgatcctgaaagaaattttcattgcttttatcagttat
 gtgcgtctggaaaggatgcagaactgtacaagcttggtcaccaggagggtttccattactgaataaaagtaagacatatgaattggaagga
 acaataatgaagacgagtactggaagacaaaagggaatggatattgtcggaaatgacagaatgatcaggatgctatattcggatact
 ggcagccattctcatcttggaacattgagttttctcctggaaaggaaattgattctcaaaaattaaggatccaacatcaaatctcaccttcaa
 atggcagccaaactattcatgtgtgatccagatctgttggtcctaactctatgtacacgcgccattaacactctcgaaggagcaattataaaag
 cattaagactgttctgcagcagcagctaactgtgatgctctagcaagactgtatatgcaagactatttgactggctgttgagaatattaacaaat
 ctattggtcaagatgttgattccaaagtacaaattggtatcttgatattatggtttgagagcttcaaaaataacagtttgagcaattctgcatc
 aacttcgccaatgaaaagcttcagcaacatttcaatgagaacccgattggaattattgactgctggatgaagcttgcatttccaaaatctac
 tcatgagaccttgcacaaagatgttcaggaattctcttcgcatcataggcttgagaagacaaaattctcggagacagattttactatctccca
 ttatgctgggaagggtgacgtatcaaacagaatcatttttgagaagaatagagattatgttgagagcactgcaatctttatcttcgtaag
 atgccctttagtttctggacttttggtaacattaccagaagaatcttgagatcgtcgtacaagttctcatcagttgctcaagatttaagcaacaac
 ttaagccctcatggaacccctcaactcgacagagcctcactacgtacgtcgtggaagcctaactctgttaatcagccctcagattttgaaaa
 tcagagtgtcctgcatcaactacgtcgcggaggtgttttagaggctgttcgcatcagcttctgctggctatcctaccagaaggacttatgctgaat
 ttgttgaccgggttgggtgttctggttcccaggtgatgcttggaaagttagatgagaggcattgacaaagggaattcttgaaaagatgaagctt
 gataatttcaacttgtagcactaaggtttccttagagctggtcaaatcgcaattttagatatgctcgtgctgaagttttgagaacgctgca
 cgccacattcaaggtcgttcagaacattcattacacgcaaagagttgtgaaaacaaggagggttcaattctatacaagcatattgcagag
 gctgtttggcaaggagaagtatatgtttaaagagaacagcggctgtatattgttcaaaaatatgtcaggagatggcggcttcacgga
 cttaccagcaatctcattcagcagctcttcttattcaatcctgtattcaggattatcgtcgcctatttctcagttatcaaaagcagaaagct
 gcttgggtatcagctctgttggaagaaacggaaggttatcttcttccaacaatacagggaagccactgtggcaattcagttgtcttggga
 gacagaaagttgcaagaagagaactaaggagactcaaaatggctgcaaatgaagcaggtgcactgctgaggcgaagaataaacttgag
 aaaaagttgtagatcttactctgagactaactctggaaggagactgcgggctgctggtgaggaagcgaagtcagtagaaatattaaagc
 gcgacaaactgatagaatcattaagtccaatgtgctgctgccaagtcggctgctcaaaagcgaacacgacaaaaatctgctactccagag
 gcagttggatgattcattgagagagataactatgttgcggagtagcaagattatgacagcagaagcagaaggagaaactccaatctgaag
 aacttagttgaatcattgtcaaagaataattcatcattgaatgaactcacctcagctcgtaaaggtatgtatgctacgatgaagaaattgaa
 agacgtagagggaatgcaaccatctccagcaaaatttgacaaattgcaggaaaaactacaacatggaatgaaaatcatgttctta
 gcaaaaaggcattaaacatgtctcgttgaacaatatgccatgactacaaaggcttttctcagaatctgctacaccgattgggcttccaaa
 tggcgagcagaagcacggatatgaaacaccaccaccagcaaaatatctcgttcacttccacagagtttaactagatcaaggagaaccagg
 atgcctgttgaaggcaggagggaatcatgaaatcttattacgatcataaaggaaaatctaggattcaaggatggttaaccagtcaccgc
 atgcatatctacagttgccttttactggtgctgtttgaatctgagaggactgctattttgatcatgtcattgaagccataaataatgttctcaa
 gggggaagagggtgacggtagattaccttattggtgtccaataaccttcttattgctatgccttctgcagaagaattttaggtcaaatgattatt
 tgtacacctcccgcaggtctggtggaactctagggttggtagaagatagtgaacactcagatctccttcaaaagcttatgggacgca
 gtgataatcttgacaagtggatgctcggatccagccatatttcaaacagcaactaactgcgtgtgttgaagatttttgggcaacttagg
 gataatctgaaaaggaaatatcaccactcttagtctgctcattcaggctccaaaatcatcacgtgcacagcctggaaggaacaccaaataca
 cctgggattggtgctcaaccacatcaaaactccattgggacaattatgcaaatcttctgacctgcttatggacactttacatgaaaattatgtg
 ccatcattctttatagtaaaccttatcactcagctatttctttatcaatatacagcttttaacagccttcttctcggcgtgaatgctgtacttctcc
 aatgggggaatgtgcaagccggctgtcattgtcggagaatggattactgatccacggatgagttgcaggaacatctatgcatgagcta
 aattatatcagacaagctgtcggatttttggatcatatcaaaaaaggagaagaagcttgaggagattaggaacgaactttgccgaacttg
 agtgtagccaaatatacaggatatgtcaatgtactgggacgacaaatacaataccaaggaatatcaaatgaggtgtttctgcaatgagg
 gaggaagtaacaaagatactcagaatctcgtatcaaatctttttattagatgacgacttatgtataccgtttctacagaggatttgcattgc
 catccctgcaatagattatgtagatatagaactccagaatctcttcacactatgcatcagtagctcctactcaagcatcatgatccacagcc
 tgtctaa

>23363 OsPN23363

caacggaggagaaaggagaaagagaggaggaggaagagatggcgagccaggggagcgcgggagggggagcggcggttctccgg
 ggtcggcgtcaccatcaccaccgcgccatgacggagacggaggacacatggcgtcggcaggaggaggagggtggcgccggcgt
 cggcgaggagcggaggagcagctgcagcgcacctcctcgcacatgcacgcttactcgcaggtgtcggagatgctggagcggggcg
 tgcgctgttcaagaacctcggccgacttcaggaccgcctctgctcgtatccacaaggagagggtggagagggtggaggaggagatcc

Figure 9
Page 3 of 9

gggagctgcgcgccgcgacgccgccaacgagcaggcacgctccctcctccacaacgccagctccacctcctgccaccgtccgcca
cgaccacacctga

>23390 OsPN23390

accccgacgacttgtctatgtccatacctccggaacttgcaggagcaattccacttattgataggttcaggtagagggttctgaaggcaat
gcaaaaacagatccattcagctggaaaacgtggttttttcgaagaaatctgtggaccacatgttcgtgaaaagttcaccttgaagatatgt
tgtgttccaaaaggatcctattccaacgtcattgctaaaaataagtagtgaccttgtaagccgttcaattaagtattccatgtcatactgaagta
catgggcatcgattccccagcaataataagttggatgaagaattgaacttgttgc aaagctttacaagcatacttgaacgttctgaactga
gagatgagctctttgcacagatttcaaaacaaacacgtaacaatcccagaggggttggtaataagagcttgggagcttattgtattatgtgc
gtcatccatgccaccaagcaagatattggagcatacttgtctgagatgttactacattgtctatggcgctacaactgattctgatgttcgagt
tctagcacttaacacactaaatgcactgaaacgctcagtaaggcagggccctagggttacaattcctgcacgggaggagatagaggctcttt
atccagccgtaagcttacaacaattgttttttctggatgaaacttttgaagagatcacatatgacatggcaacaactgttctgatgtgtgag
gaacttgcgtgcattatcaaaacttcagtatattctagttcagccttttgaatgccgaaagggttgaatggatccaagcttcagatgttggcaa
tgaggagtagattgggttagatgataacaaatataattggtgacctcctatctgaattcaaggcagccaaggatcggaacaaaggagaaatttta
cactgcaaaactgtatttaaaaaacgcctctccgtgagtcggatgaggctataactgatccaatgtttgttcagctatcttatgtccagttacaac
atgattatattttaggaaactatccagtaggtaggatgatgtctgcacagctatctgctctgcaaatattggtcgaattggccttgttgacaatcc
tgagtcttgtgtgaatggatatctcttttagagagatttctaccagacaagttgcaattacaagagcaaaagcgagactgggagcttgacattg
tctcacgctatcagttatggaacatctgtcaaaagatgatgcaaggcaacagtttctcagaatcttaaggactcttccatattgggaactcagttt
tcttcagtgctccggaagattgatgatccaataggacttttacctgggaagaatcattctgggaataaacaacagaggggttcatttttccgtccg
gttcccaaggaaatattctatctgcagaactgagagatacatgcaatttgggagcagcaataaccgctgttttcttaaaatgagagtcgctgg
tgttctcatatctttagtttgaactaagcagggcgaggaatatgtgtagcacttcagacgcataatgatgtgatgtctccgcagatattc
aaaagcacgctctgtactagcgcggttttctcaaaacgatgttttctcaacataataaccaccgaatattgaaatatatgagaaacgtgtccaa
gaactgtcaaaagcagttgaggaatctgaaaggaaagcagatctgttgaatgaggagttacagaagaagacaaaaaagaaagagatatg
caaaaagaattagaaggtctgagggataccttgaatctgaacggcaaaagcattaaagaagtaacaatgatcttgataaactaaatccctt
tgtgatgaaaaggactcctcttgcagggttctactgatggagaaaactagactggagaccagattaaaaagtggtcagggccaagaaagca
gtaatagaacaggagtagatcaggaaatcattttgagagagatactctccaactgtaggcactgtcaataatagcattgagatgttagcaaaagct
tgaggaggagttaaaatcctgtaagaaggagcttgatgcaccaaggaattatcaaaagaattaacaatggaaaataatctgcttgaccagaa
ggttcaaaaggcttgaagagcaaaaaagtgaaagagaaagtaatatggaaagagtttatgaggatgaatgttgaagctgaaatctctgattg
ctgaattggagcaaaaacttgaaagcagaacacgttcctgaatgttactgaatcactcttgccttaagaaatgctgaggtggatacattgc
aaaacagctcacaagaacttgacgagttacgagagttcaaaagcgatgtggacagaaaaaacacagactgctgagattcttaaaaggca
aggagcacaattgattgagctgaaaatctttataagcaagagcaggttctgagaaagcgttattataacacaattgaagatatgaagggaaa
aataagagtttttgcgtctgcgtcctctaaatgataaggagctcattgaaaaggacaagaatattgttgcagccctgatgagttacagttgc
acatccatggaaagatgacaagtcaaaacaacatatatatgacctgtttttagtctaacactaccaagaagaagtatttgaggacacgaa
gtatctagtccaatggccgttgatggatataatgtttgtatattgcttatgggcaaaactggttctggaaaaactttacaatttatggatctgaaaa
caatcctggtcttaccceaagggttacctctgaactttttaggtgataaagcgtgatggtcacaaatattcattttcttgaaggcatatatggtt
gagctttatcaggataatcttggacctgttggccaaaaacgcaacacaccaaaaaattgaaataaaaaaggattccaagggggtgtt
actgttgaaaatgtgacagttgtgaacatttcaagtttgaagaactgagggtctataattttaagaggttccgagagaagacacaccgctggaa
caaatatgaatgttgagagctcaaggctcatttaattcttcaatcattattgaaagtaaccaacctccagactcaatcttatgcaagaggaaagc
taagtttctggaccttgcgttctgagagggtgaaaaagtcggctcagcaggaaaaaactgaaagaagctcaagcatcaataagctt
ctttctgcattggctgatgtgattggggttattctctgatggacaacatatacctatcgggaaccataagttgacaatgctgatgagtgattctct
tggaggcaatcgcaaaaccttgatgtttgtgaatgtctaccagcagagtcacacctggaggagacttacaattcactcatgtatgcttcacga
gtgcgttgcatgtcaatgatacagcaagcatgttccccaaaggaaatcatgaggttgaagaagtttaattgcttattggaaggagcaagct
ggcaagcgtagtgaagacgacgacctggagggaatacagggaagagaggacaccaaaaggagaagcagataatcgcttgactagctga
>23416 OsATPF

atgaaaaatgaaccattcttctgttttttagctcactggccatccgctgggagtttctgggcttaataccgatatttttagcaacaatctaataaa
tctaactgtatgggttggtgattgatttttggaaaggagtggttaaaagatttattagataatcgaaaacagaggatcttgagtactattcga
aatcggagaagaattgcgtagagggaaccattgagcagctcgaaaaagctgaattcgattacagaaagtcgaactagaagcggatgagtatc

Figure 9
Page 4 of 9

gaatgaatggatactctgagatagaacgagaaaaagcaaatgttattaatgccacttctattagtttgaacaattagaaaagtctaaaaacga
 aacctttatttgaaaaaacaaaggcgatgaatcaggtccgacaacgggtttccaacaagccgtacaaggagctctaggaactctgaata
 gttgttgaataccgagttacatttccgtacgattcgtgctaattagcattctcggggccatggaatga

>23484 OsPN23484

atggatccaatcggcgaccatcgcgctgctccggcggtcgggtgaagcggcgccctccgcgcgatcggcgagctgtcgcggaagg
 cgtggggaggggaggcgccggagctgatccggcggtggaggagctggaggaggcgccggcgccggtcggggcgaaaaggagg
 ccgcggaggaggcgccggagctgcaggccgagttggatgcggagcggcgctcggcggaacggcgaccagcgaggccatgct
 catgatcgagcggctgcagcgggagaaggcgccgcgcagatggaggcccgcagttccggcggtacgccaggggccgcgaggatc
 gcgagcgtgaggtacaggaagagctgcctcgtctccgacctcggcgctcctaccactcccgcctccagttccacgggatcgacctg
 attccttctccgacgacggcgaggaggagcagcacgacgaggaagatggcgaagaggtagaacagatcgacacggcgcgctacaga
 cagatggaagcagcggcgccgactcaatcgggtgcatgcaggtgaagccatggtcgacgatgatgaggaagagcaattcacctgtg
 gagaaggagttcagttacacgggtggtgaggtgcgcgagctcgacgacgaaggatccggggcggtgtgttggggagttcatggg
 ggaggggcaatcgagctgcagggggattgtacgccagggtggaggctctggaggcgacaggcgggcaatgcggagggaaatcgag
 cattcgagcagagagagctcagctggtgatggcgaggcgatggcgcgccggtgtgccgggaggtggtgtgaacagaaagcca
 agaaagtggctgcctaccacgaagcttctcggtgctaggggttgcgaagtgggtgctctcgataatcttttgagaacagaagttccacag
 ccaggtatccctcggctgtcaactacgttcttggcttctactgctgctcgacagatccaccatgttgagtcctggcgctcgtcctagc
 ccacggcaatga

>25358 OsAAK39589

atgaaagagcccgcgaggtgaggacggcgccggggcgccgcggcgatgggggtgaggtgggggtggaggaggaggaggagcc
 gccgcgatcggcgacggtgaagcaggaggaggcggaacgcggtgctcggggcgagggggtcccgccgttcgccatcggggagctca
 aggaggaccacgaggtcgcggccgggagcggcggtgaaggcgcgctcggggagaggaaacgggatcgccggatgccaggg
 ttcgtcgtacagccaagagagtatgcagcagtttcatccatcatgatgttgaatggacttaataatagtgtactggagttgacgaggaa
 ggccgttctcgcagaggttcttttctgtccaaaagatacataagcgcaatagaagaaatcatgatgacctgatcatattataattg
 ggccctgttctccaggaggtgcagacaatgtatgataaattctagtctcgaagatgcattgcttggaggaggttgcaaaaagtagtct
 gaagctacccgtctttgcccaactctttatgatcatattataactgggctattgctattgctgatcgggctaaaaatcgagggtcgtactaaagaa
 gctgaagagctctggaagcaggcaatactgaattatgagaaggctgtccagctaaattggaatagtcgcgagggtcttaataattgggtctt
 ggactacaggaactgagtgaattgtccggctcgtgagaagcaaacatcataaaacagctataagtaagttcagctgcaattcaactg
 caatttgatttccatcgggcaatatacaatcttgggacagctctgtatggttagctgaggatacaatgagggtctgggaaggcaggtgtgtccg
 ctacgagttttacagtcagttctgtatctatgttgcagctgccatgcactgaagccaaattactcgggtctaccgcagtgcttgcgattagctc
 gttcaatgctgccttggccatctcctcaagtggtgatatgttgcctcctcagaaaatagtgccattgcaccacacaaagaatgggagaggtc
 acagtttgtttgaacctgaagaactccagcaggtcaatgcctctgaccagcctccatcacatcacctgggcatgtggacagcggcagaa
 agcttttcaggatagttgttcagacattgttctgtgtcagcatgcgctgatctaaccctgccacccgggtgctggccttcgcacgatactattc
 atggccctagattcttgggtgctgataactgggagaccattgacagttggctagacgctatacgcctagctacaccatatttgaagaggag
 gattgatgtccttgcgtgggattattactggctga

>25381 OsAAK20062

atggtgatcaacctagaattcatcaggcgatcgttacggcgatgagatccttctactggatcctctcacgatcgtgatccctttgttga
 gcaattgacacatcatctcctctcaagaacctgggtgtgtgggaatggtaacctgggtggtgatgacctggggaaaagcatgatgactcgc
 atggagatcaggtgccccgccttaacgaggccaccggagcagagcagagttgccattgaattccaggtgctggagctgcactcgaga
 ccgtgtgctcatcattcgacgttaacgtgtctggtcttgaaggcgccgactccggtacttgaggaaactgaccaagaatgtcagcacaagg
 aatctcgatcgtgtgcgaactcicaagagtgtatcttaccggttgccttcccattgtgcagaagggtcagagatgaaatagaacatcttctggatg
 ataataagacatggcacatctgtatctaacaagggaagcaattacaaaaccagcaagttgaggctttaatatcatcggtgcttccaatagcat
 tgttctggaggaaacagttgtccagggtgaacaatagtttccggcgagcgtgagcatcgctacaagcatgatttggataatgacgtgga
 agacctagaatgttgcctgagggttacttcatgcaattggatggaattcgaacagaatttgcggttcgagagtattgatgatacagaag
 actatgtcaattatcaacttgacaaccagcgtaatgaacttattcagcttcagcttactgactattgcaccttcggcatagctgtcaacacct
 catagctggggcttttgcgatgaacatccagagcaaatgtacagcattgatggtagcttcttttggccatttgttgggggacacctcatcgg
 gctgcttcatgatctgcacgtcttattatggtatgcccggtggaagaagttgctcgggtccttga

Figure 9
 Page 5 of 9

>26210 OsAAK38489

atgtcgtcgcgcgcgcgcgtcgtcagcttctggaaagactttgatctggagaaggaaagaggtggttggatgagcagggtttaagatt
gcagagaatcaagagaccagccagaaaaacagacgccggcttgcgtgagagcactagggtttcaagaaggcgtcttcagatgataagctt
agttattcaattcttctgaagagttatcaagaggaaagtgataatctcacaaagagagcaaaatttggagaaaaatgcatttctcaacatctac
cagaagctctatgaagcccctgatccataccagcttctgtctatggctgagcaagatcagaaactatctgaactggagactgaaaatcgg
aagatgaaacttgagcttgaagagtaccgggcagaagccgccacctaagaaccaacaagcaacaattaggcgacttgaggagcgaaa
ccgccaactagaacaacagatggaggaaaaagtcagagagatggttgaatgaagcagaggagtttgcagaagatagccagaaaactc
ttgaagccctgaagacagggagcgggcatcgcaggaccagctaagacaagccacagaaagtgtaaagaatatgcagaagctacatgaa
tcagcacaagtcattgttgcgttctactcaatcagaggaggaccgggtgcgaaggaaactgaagtcaattcttctgttgatgaagt
gaacggggcccaagcacgtttagttagcttgagagagaaaaaggtgatttacatctcagttgcaacaacaatgaagatgcaaccaatag
tagtgactacgtggactcaagtgtatatacttgagagttccttgaatgccaaggagaagatcatctcagagctgaatgcagaactgcgtagtatt
gaaaatactttatcttctgagagagaaacacatgttaatgaactgaaaaagttgactgcttgcgtcagtgagaaggaaaatgctctgacagagt
tgaagaaagaactcaagaagaccaacaagaagactagtagatgatctcaagaaaaaggttcagatttgcaggctgtagggtacaactc
cattgaagctgaagattgggagctagcaacaaatggtgaagaaatgagcaagttggaagcttgccttctgataagaaccgtaaaatggagc
atgaactaacacagctgaaggtcaaaatatcagagaagtcacatcttgcaggaggctgaaaaaaagattgctgagctaaactgcaaggct
gaagagcagcagaagttgatttgaactcaggatgacatactaaagggtatagttcaactgtagggaggacttcacttctgaatgattgg
gatcttcaagaaatggctcaaatgaagtagcagagggtactgatccaaggcctgcaccacaagaccaagaccaagttccatgcttaagg
tcatttgaatcagagggatcgtttccgtacacgtctacgtgaaactgaggaggaactaagaagactaaaagagaagatgaaatgctagtgtg
tagaattggaaaaaactaaagcagataatgtcaactgtatgggaagattcgtacgtgcaggactacagccatgagaagattgttccagag
gaccaaagaagatgcagaagatgttgaagtggttcttcagatgttgagacgaagtacaagaaaatgtatgaggatgacataaatcctttg
ctgcttttcaagaaggaaaaaggtacacgggtacaaggaacttggtttaagagacaaaatcactcttagcagtgagcgttttctccttgtaa
caaatatgcccggacattatattctctacactattggattacatctccttgtattacattgctgtacagaatgtcagctttgagctatctcagacg
tctgaatacgttttctgtagataaaaaatttccagatatggagacggggtggatgggaaatgacaggatgaggcgtggtcgtgctttagacc
gttattatgtaaactgttgcatactaa

>26688 OsPN26688

atggcttctcttctcctccttggccttggcgccatcttccaaagcggctgccccctctcccgcctcgcgcccgccgtccgccgcgcgcc
caccaggcgccgcgccgtgccaccaagatctcctgcacggatgggaccccagggcgctcctcggcccgccgagggcgccacat
cgtgcgcctcaggttccgccgccgctcgcagaggactccgacgcccgagggccttcgagcgccaggtccgcgagggagcagagcg
acgccgccaggagcgcgagggcggggtcatcctgcacaggacgccggcctcgtcaggttctcctcgcacaggagcgcgcgagatc
gaggtagagatcggcaggctccgccccaggctcaaccagcccttctcgcactacatccagcgcgagatcgccagatcaagttctaatca
cccgaacagcggaaatggaggaccgattgatcagattggaagcagatgcaaaaggttctgcttgaggagtgaggcctatgacaagttgc
aaaatgaccttgcagcgcgaaagaacgcctcacaagatcctgcaatcaagcgacaaaaaatcgacacttctgaaatggtgaacggaa
tgagctcaacatgtccatactaacactcctcgatgaaaatatagcaagcgccaagaccaacaatcaggaagaagctgtggcttcatggaga
atgttcgctcatcctcgtgaaatatattacagtatga

>29882 OsPN29882

cacgcgtccgcaaatgagcacattcttgcctatggaaaaggaggtgaaaaattgcaggctcagctgaagcaagaatcattgctaaggcagc
aggagcaacagaaacttctgaagagtccttattaaggcaacaagagcaacaaaagctaactggagaacagctcctatgctgcttcttgggt
gcggaaaagaaagatttgaagaaaaaattgctgccttgacgaagaagcatcagatgaagcttctgaatttgcacgcaaggcattttc
aatggaagatagggaacttgaagccaattgcatgataggttgcgttgagaggctagaggcgagccgcaaaaagttgtaattgg
agattgattctcaatcagaaatagaaaagctgtttagggagaactcagccttgcctacttcatccaagaagctgtagctgttacaatgcaa
tgggaaaaccaggtaaaagattgtctcaagcaaaatgaagagctacgtctcacttagagaagtttaagacttgagcaagctactctattgaaa
acaagtaataccactatccagccagatgggcagaacgagaccagcatttcattcccgcagagtttgaactgagaatctctccctaaaggat
cagctcatcaaaagacagagcagatccgaggggtgtcggcagaataatgaaacttgcagctgaactcagaaaagcgggtgcaagcgca
gaataacctgcacgcctatatagacctgttgaagggcatagagagcaacttaataaaaatgaaacaagaacatatgcaacgactccagt
aa

>29942 OsPN29942

Figure 9
Page 6 of 9

gagcatgaatcagatgaagagaagaatgaagttgatactagcaatgaaatcaaagagattgtcattgatgacatggagccaaaagtgtcca
 ggctgggtttctttatgtacagggacaatcttgggtgatgatgagttgtctgcatcaagctcggactgctctatctttgatactttagctggga
 aattattggctgctgctgacagatgatgagttgccaagactacgggtattgtgtgaatcttacttgtgcaagcatatttctgaaactctgtggcaac
 taccctagcattggctgaccgacatcatgctatggagcttaaatctgtttgcctaaaatttgcctgctgaaaatcttcagctgtaatccggaccga
 cgggttgattacctcaaagacaactgccccgcgtgcaatcgagatactgaggacggcggcggtgtgaggaggaatgcagcagcg
 gagggaagagccagagcgtgtgggggcagctctcgatggcggcgataccagcggcgaggggtgaggccaagggtctga
 >29956 OsPN29956

tttgagattgaatgggagctgattgatgaaagaaagaggagctacaaaaggaagcgatcagaattgctgaagaacgaagagcaataact
 gaggatctgaagaatgaatctgatatcaacacaggagaaggataatctccgtgttcagttcaaaagtaattcagaacactctctctgtaac
 acaaagagttcatgagtaagatgcagcaagaacatgcaagttggctgagtaagattcaacaagaaaggcaagatctgaagagagacattg
 atatccagaggggtgaattgctaaattctgtaaggcaaggcagatggaaatgattcttatttggggaaaggagggaagagtttgagcag
 aaaaaggccaaggaactgaacacatcaattctcagaaggagatgatcaacacaaaattagaacatgttgacgttgtaattgcagaaacttaa
 ggatgagagaaaagaagctactttggaacgtgagaggagagagcaagagttgtctgagataaaaggcactattgaagccttgaataatcaa
 cgggagaagctgcaagagcaaaagaaaactattacattcagaccgagaagcaattacagtgcataattcaacaacttaattgtgttggaagaact
 gaaaattgatttctgaaaataagcaactgtctttgttacaacatgataagtcaagccttg
 gaagtgatataaatgtgaaagacaatcatcatgataactccattctca
 ccaaagcaacgttttgaaggaactagacctttctccagtctcaacaccgatttcttgggttcgaaaatgtgctcaggtgatattcaaacggtc
 tccagagaagagcgctagccatgatcaattgtccagaatgggtgtccaaagaaagttggagactctgtggatgtgaagatgtgaatttga
 ctttgcaaaagttgggtcaaaagaggcttaatactttgttctgtgaccaaactgaagtttagagccaaaacgaaagcacagaaggagtagt
 attcagaaaagttaatggaggggaaatcactccaactgcctgtcagctctagaagagaatgtctcaagaatgaacatgatgaagcccc
 >29957 OsPN29957

atggcgaaagtcgagcgcggacgacgccgagctgcggcgcgctgcgcgcagggcgtggcggttcgggggcgcgcggcgaggagg
 tgtcctctccatccgcgtcgccaaggggaggggcatcttcgagaagctcggccgcctcgccaagccccgcgtctcgcgtcaccgttaa
 acaatcgacaaaaggtgagggcagccaaagctttctccgagattgaagtattcatccggcgagctactgagccagccaagctttacaagtt
 gaaacatctcacaaggttgaggttatttcgaatgacctagtggttcacgtttgtattgggttgataaccitagaagtcagagtggtcccc
 tccacaatggacaatgcgcaacattgatgacagaaaccgctactttttctatctgacctgtgcaaagagatacttagctatcttccaaaag
 tegtgtgaattgatttggagctagccctctgggccaaggaaaatacagtgactctggacaaccaatcgagcacccaagatggacaag
 aaaaatctgtcacaactcaaactgagaggaaagtaactgtaactgttgaatatgatcttgggtcccaagcaaaaggaatgaagaggaagacat
 ggaggcacttctgacacatgatcatggcataggtgaagcagatgctttctctgagagattgaaacaggaactgtgtgcttggaaagcagc
 caatgtctatcaactactgcagagtgaaactttaatagatgaggttttgagggtctggatgctgctagtgcacccgtagatgatatggatgag
 ggttacgcaattttaatgaagctcaggcatatgagagaagatattgcatcgattgaatcacgcaacaatggcttggagatgcagtcctgtaa
 taacaaaggactcgttgaagaactagaaaagttgcttgatcgcttgcgaattccacaggag
 >29958 OsPN29958

tcattgcagaatgaagtcagtgccctagagaacaaaccttatcccttgccaatgattgttgcaatcaaataagctcaggatggaggaaaat
 gcattatcaactcaggttttgaagaccaatatgagatcgagtggtgatcagaatacagtagcgacagtaaaagacatggagctgcagaaatt
 gcatgggacctcaaaagcactccagaaggtggttacggatacagccgtcttcttgatcaagagagggcttgatttcaatgccaatctgcaag
 aagcgaggaagcagattgaggtgctgaagctgaaggagattttggatgatgacttgattgaaatgaattatgagcaaatgctgaaagacatt
 cagcttgacctcatccaaatttcttctgtaataaaaccgggtcccttggtcaggccaataaaaactgtagcacaggcaaatgagaagatgcttg
 actctcatggcatcgttggagctagcagtagccatgtacgcaatgatttgagaccgccgaaagtgaagtcgttggaggggacaattacaaa
 aggcctccctctgaactgatggctgtgaaagaactcagcattgacaagcaagagttaccaaggtaaccaccagggagccgcaccaagag
 tggaaagaacaagtaatcgaagattggcttctgatgcacaaaggctcaatgcctccaatccagtattcaggagctcaaaacaaacacag
 aggcacagaaagggtcgcagctcgagagcgttaggtaccaataagagaagctgagggttcacactcagctgatcgaacggta
 aactgtccaagaaggctgaggagttcacatctgaagatggcttgatggggacaacattgacttgaggagcagacaccagcgcaagatcat
 g
 >29961 OsPN29961

Figure 9
 Page 7 of 9

gatagtgcctatggaccacaagtatgggcaaaaagatgggtgctcctgatgagggttctgtaggtgttcctggtaggaaaaccaaagagacagt
 aactgctgctgatgcaattatgatgcacttgacactgccgaggaagaagtaaaacgccttgatcagcaccaggaaggacaaaataatggg
 aatggactacattcaaccaatgttataatgcaagggcagtcaccatcggtatgtttgaatgtagtttcaaagtgtcgtccaaatgacttgg
 aacaggctctccttcattgccatttcagatgcacttaagatcatgtcttacttaaaaggagtgggtctatggttcctttgaaggttgagcttgttga
 gggtttgcctagtgtgcttcaaacacaccacagccaattaactacgacaccatctgcaagatcgatacttacagaactgaaaggcattcttta
 cagtagagtaaaaggaatgcaaggatgctataggttcaaccttgctgcaatggatcatataaaagaattgttgccatgagatcagatgctcc
 ctccgagatgcaagagcgaagctgatggagatcaggcaggaacaatcaagacgatcgataggtcagatggggctgaaaaaagga
 aagaaaaagcgaagaccatccgggtgagagctga
 >29965 OsPN29965
 ctactagaggaaagcacgtacgcgaaaggcttggttcggctcgggagttgaattgaaagcgttgctggaagaagtcaccaagctcatg
 aacaaaaacgagaagccttgcttctgagttggcatcggtgagaagccaacccgcgtagagccaacagtggaactgagaggtactcgag
 ggatagcatcagtagacgacatgagccagctccaagaagagacaacaacgcagggtacgagagagaaaaagcattagaagctgtgctta
 tggagaaggaacaaaaggaagcagaactccaaggagaattgaggagtcgaagcaaaaggaagccttcttgaaagtgaacttgcaaac
 atgtgggttctagtggcaaaactgaaaaagctcaaggccatgatcttgaggatttgacaccaaatatattggctcgtaa
 >29966 OsPN29966
 gttactgactctgttatgaagggaactataagtgtgtagagcaagaatcagctcgccagattgcatctaaggatgcagagatagcattcttga
 acgaaaagctgcatcaattcagaaatagtgccctgagttatctgagggccgagataaactctatgaagaaattataaccttcgccaacagct
 tgtgactcttctaagtcaacttgaattcagtgaggggcttcagtatcacactataacaacttgaaggcgcagaagatgaaagtaagcata
 gaggggaatgaaaaatcctctaaagatggtataactaaggaaaatggtctaaaggctccaatgaagatatatttattgatccaacagttctgaag
 cacatggatagggatgagttgtagccatttcaataaaatgatgaatcaaatgaaacggcagcatgattcgaacttgaagagaagacgga
 agagatatattgactaaagcgggagaacctaaaaaagaaggacctaacccctggcatttacgaataataaagaatttgagctcatgagga
 agaaaatttggaagttattacaaaattggatgaggtccttgaggagaataaaagaacctccgcatcaagagtgtatcttctggtcaaca
 agac
 >29967 OsPN29967
 tctctctctctctctctctccctccacccacgcgccccgcacccaccacgcgcccgcgcccctaacccctaacccctaacccctact
 cctctcgtcgccatgtccgagcccgcgtgccccaccggccatgcccgaggacgcccgcgcccgcagccgcagccggagccggc
 ggttcagcgggggaggaggccgcccgtcttgagagggaaggaggagttgctgccgggtggaggagaagatctcggaaattagatgaa
 tcacagtcacagcttatggggagacttcggggactcaaggaggattgctgaattggaggagcagtttagatacacaagtacaaaatacaa
 aagcgaactctctgatataaaaactgcgctgaatagcgaattgaacagctgagatcagatttcaagaattgaggaccacccttaagaagc
 agcaggaagatgtatcaaatgtttgaagaatttggggctacaagatgctactgacaatgatgggaacaaaggaagt
 >29968 OsPN29968
 atggagaagtcgccactggcgtggttcaggttgctgtaacaatgaagacgtcgttgccataaagcagatgcagcatctcatactgggac
 gattgcaagatagtaatgcagttctcacacattttaatgaatactctgaacaatgctttgcggaggtgtctaatgacttgggtgtaaaactcgtct
 tctcaagtcctgaaggctgatcttgaccatatttctgaagttgagaggcatgaaatcgaggtagcggaacatatccagatgcttttcta
 ctggcgcaatggcagagacgatggaccaagaccggaccttgaaagtcctcttgattaa
 >29969 OsPN29969
 atggagaagtcgccgcccggagacagccgcccggcggcggagggtggaggcgcggttcaggctgctggtcgacaccggggacatcgg
 cgccatcaggcagcgcagcacctcatactgggacggttgcaagatagcaatgcagttctcacacatttgaacgagtactctgaacaatgctt
 tgcggagggtgtctaatgactttagtaaaactcgtcttcaaatccatgaaggacgatcttgaccatatttctgaagctgagaagcatgaa
 atcgaggctagcggcaacatatccagatgctttcccgatgtgcaatggcaaaacgatggaccaagaccggaccttgaaagtccgctt
 gattaa
 >29970 OsPN29970
 tcgtcctccgaccaggctgcagaatttactgatatggaaggcgaatcatcagcagttacttaccatttctgctctgacttcaaccacaccaa
 atgaattggagatgactaacaagaattccaatgtggttaggcggaatgactcacagcaacagtatgccacattgactgctgcaaaagatgga
 aatacaaaagtttacccttgaattccgagcacttgaagtgtgcttgagtcagcctgtagaagcctagaagaggaaacgtgactctggag
 caagaggcgtatccggcacttgatgagctgacttcaagataagcacactaaatcttgagcgagttaggcaattaagagccgctggttgc

Figure 9
Page 8 of 9

aatctctggctggtgcagaaggtgagggatgaacttgagcatttggtgatgatgaaatggacatggctgaaatgtattgacagaaaagctt
actcgacaaga

>30848 OsPN30848

ccacgcgtccggccttggaactttgagctttcaagatgccaaggaggacagataatgctgcttctgccaattcagttgaaccagaaaaatca
gaagaatgtttggagtttgatgatgaggaggaggaggtagaagaggaggaaattgaatatgaagaaattgaggaggagatagaagag
gaggaggtagaggaggtgaagatgctgaggaggaagttgaggaggtatgaggaggaagatgaggaggaagagaggaatctgat
gaaactgaaggtgtaagtaaaactaaaggtgtcaccagaaggaatgtactgaaaagggaagcatgctgagcttctgcttctctcccat
ggttctgaagtatatgttgaggcatctccagtgtatcttctgaagatctgaagagactatgtgaaccagttggagaagtgtggaagtga
gaatgatgaggagaaaggacgatagcaggggatatgctttgttaatttagaacaaaagggttggcattaaaggctgtcaagaattgaaca
atgcaaaactgaagggaagcggataagggttctcctgcaggctaagaacaagcttttattgaaatgtacccatagttggacagatg
atgatttcagaaagggttgagggaagttggtccaggagtattaaagctgatctcatgaaggtctcaagtgcacaaatcgcaatcggggtatg
gtttgtgaataactacaacatgcatgtgcagagtatgcaaggcaggagatgtctcccaacattcaaatgctcctacagtc
agctgggcagatcctaagaataacgactcggcctctacttctcaggttaaatctgtatgtgaaaaatctgcccaagaatgttactcaagcac
agctgaaaaggctatttgagcaccatggtgaaattgaaaagggtgttcttctcctcaagaggaggtcatgataataggtatggtttgtcact
ttaaggacagatccatggccatgagggtctgcagaacacagagagatgatgcttgatggtcaggtccttgattgctcacttgcgaaacct
cctgctgctgataagaaggatgatagagtaccactacatgttcaaatggagctccattgctccgagttatctccacttggatatggtatcat
gtcagttaccaggtgcctatggtgctgctcctgtagtactgcacagcctatgtgtatgtccaagagctcctccaggtgcagcaatggttc
caatgatgttaccgaggtgctgctcgtatgtgtacaacagcctggtggacagttgccgctggcttcgccgccgccgagcaagctgga
catcgtagcggcagtgaggacgtcatggcggcagt

>30854 OsPN30854

atgcagggggaggtggatcagccgatgcagatggtgctgagggtgaagcaccgctcgtcgtggtggcgggcgggcgggcgggcgggga
ggaggaggccggcgaggcgtcgtcaggtcggcgctgctggtgttcaaggccaaggaggagcagatcgagaggaagaagatggagg
tgaggagaaaggtgttcgcagctcggcgctcgtcagggaggtccaagcgcttgccttcatccgacaggaactggaagggtggc
ggatccaaccaggaaggaggtggaggtgatcaggaagaggtcagctggtgaaccgtcagctcaagcctctcggcaaaacctgcgtc
aagaaggagaaggaatacaaaagaaatcctggaggcgtacaacgagaagaacaaggagaaggcactgcttgaacaggctaattgagct
ggtgagcgaagtgagaggatgcgtatgaagaagctcagggagctgaacaagacggctgattc
actctactag

>30899 OsPN30899

gttcttctgattctttgaaacggaagacatatgatgatgagctaaggaggaggagcttttgaactacttcagacggtttcagagtgttctca
aaagaaaggaggaagtggatatttgcacaagggttagcccttctgaaggtgtgatgaaggaccttatggtttatcaagaagaatagcctgc
aagaatgtggtgacttccacctgtggatttatacaggaagagctaaatcgcaagctagatggtgtcaggactgcaatgatttcaccaagct
aaagatggggatggatgggtgagcagcttttcaaccagttctatttgggtgctgcacaagcctgaattgcctcatgcatatgtttgtgctgaa
agcatcatttcatgtcaccgagtggttcacctgtcagggaatgagatgccccgcgaacactcacaagccgagcttccatgtcaatgccag
ctgttaaacgagaatagtggcaaggaggagcacctcggcgagagggtggaggatccctaattggtgtaaacatggatggcggaatcg
atgaaggaggagttcttcagtggttcagaatgcactgcagcttgcatgtttgaatcgtttggcgcgcaaaacgagcccccttccctggaa
gtggaagcaatgccaagggtagtaatagcagtagt

Figure 9
Page 9 of 9

Figure 10

>19788 OsMADS1

atggggagggggaaggtggagctgaagcggatcgagaacaagatcagccggcaggtgacgttcgccaagcgcaggaacggcctgctc
aagaaggcctacgagctctccctctcgcgacgccgaggtcgccctcatcatcttctccggccgcccgcctcttcgagttctccagctc
atcatgcatgtacaaaaccttgagaggtaccgcagctgcaactacaactcacaggatgcagcagctccagaaaacgaaattaattaccaa
gaatacctgaagctgaaaacaagagttgaatttctcaaacacacagagaaatattcttggtgaggatttgggcccactaagcatgaagga
gctggagcagcttgagaaccagatagaagtatccctcaacaatcaggtcaagaaagaaccaagcactgcttgatcagctgttgatctga
agagcaaggagcaacagctgcaagatctcaacaagacttgaggaaaaagttacaggaaaccagtgagagaatgtgctccatgtcct
ggcaagatggtggtgggcacagcgggttctagcactgttcttgatcagcctcatcaccatcagggtcttctccaccctcaccagatcagg
gtgaccattccctgcagattgggtatcatcaccctcatgtcaccatcaccaggcctacatggaccatctgagcaatgaagcagcagacatg
gttgctcatccccaatgaacacatcccatccggtgatatga

>20072 Os000564-1102

atgtcggcacaggcggagcttccctgaggagaatgtgtacatggccaagctcgtgagcaagccgagaggtacgaggagatggtcga
attcatggagaaggtggccaagacgggtgactctgaggagctcaccgtggaggagcgcacacctctgtctgtgcatacaagaatgtgattg
gagcccgcctgctgcatggcgcatctctccattgagcagaaggaggaagccgtggtaacgaggaccgtgtcacactcatcaagg
actaccgtggcaagatcgagactgagctcaccaagatttgcagcggcattctcaagctgctgaatcccaccttgcccccttccactgccc
ctgagccaaggtcttctacctcaaaatgaagggtgactactacaggtaccttccgaatttaagaccggggctgagaggaaggatgtgct
gagaataccatggtggcatacaaggctgtcaggacattgcttggctgagctgcctctactcatccaattaggcttgggctagctcttaact
ctcagtggttctactatgagatctcaactcgcctgacgtgcttgaacctcgcgaagcaggctttgatgaggccatttcggagctggacacc
ctgagcgaggagctctacaaggacagcacttgatcatgcaactcctcgtgataacctgaccctgtggacttcaaacatttcggaggacc
cccggaaaagatcagggaagtccaagcgcgactccagcgaggggcagtaa

>20231 OsMADS45

atggggaggggtcgggtggagctgaagaggatcgagaacaagatcaaccggcaggtgacgttcgccaagcgcaggaatggcctgctc
aagaaggcgtacgagctctccgtctctgcgacgccgaggtcgccctcatcatcttctccaaccgcggaagctctacgagttctgcagca
cccagagcatgactaaaacgcttgagaagtatcagaaatgcagttacgcaggaccggaacagctgtccaaaatagagaaagtgcagaa
tgaaagctagccgcaatgaatacctcaaaactgaaggcaaggggtgaaaattacaacggactcaagaaatttctgggtgaagatcttgatt
cattaggcataaaagagctcgagagcctagagaagcagcttgattcatccctgaagcacgtcagaactacaaggacaaaacatctggtga
ccaactgacggagcttcagagaaaggaacaaatggttctgaagcaaatagatgccttaggagaaaactggaggaaagcaaccatgttcg
cgggcagcaagtgtgggagcagggtgcaacttaattggctatgaacgtcagcctgaagtgcagcagcctcttcacggcgcaatgggtt
cttccatccacttgatgtgctgtggaacccacccttcagattgggtaccctgcagagcatcatgaggcgatgaacagtgcgtgcatgaaca
cctacatgcccccatggtaccatga

>20232 OsRAP1B

atggggcggggcaaggtgcagctgaagcggatcgagaacaagatcaaccggcaggtgaccttctccaagcgcaggtcggggctgctca
agaaggcgaatgagatctccgtgctctgcgacgccgaggtcgcgctcatcatcttctccaccaagggcaagctctacgagtacgccaccg
actcatgtatggacaaaatccttgaaactgtatgagcgtactcctatgcagaaaaggtccttatttcagctgaatctgacactcagggcaactg
gtgccacgaatataggaaactgaaggctaaggttgagacaatacagaaatgtcaaaagcacctcatgggagaggatcttgaatcttgaatc
tcaaagagctgcagcagctggagcagcagctggaaaattcgttgaacafatcagatccagaaagagccaactaatgctcgagtccattaa
cgagcttcaacggaaggaaaagtcactgcaggaggagaataaggctctacagaaagaaactgggtggagaagcagaaagtcacaaaagcaa
caattgcaatgggaccagacacaacctcaacaatttctcatcctcctcatgatgagggaagcccttccaacaactaatatcagtaact
accctgcagcagctggcgaaagatagaggatgtaccagcaggggcagccacagcatgttcgattgggctgccaccatggatgctgagc
cacatcaacggctaa

>20233 OsMADS6

atggggagggggaagagttgagctgaagcgcacgagaacaagatcaacaggcaggtcaccttctccaagcgcgcaacggcctcctca
agaaggcctacgagcttccgttctcgcgacgccgaggtcgcgctcatcatcttctccagccgcccgaagctctacgagttcggcagcgc
cggcatacaaaagactttagaaaggtaccaacattgtgtacatgtctcaagattccaacaatgcacttctgaaactcagagttggtaccat

Figure 10

gaaatgtcaaagttgaaagcaaaattgaagctttgcagcgcactcaaaggcacttgcttggggaggatcttgaccactcagcgtcaaaga
attgcagcagctggagaaacagcttgaatgtcactatcacaggcgagacagagaaagacgcaactgatgatggaacaggtggaggaac
ttcgagaaaggagcgtcagctgggtgaaattaataggcaactcaagcacaagctcgaggtgaaggtccaccagcaactacagagcca
tgcagcaagcctcctgggctcagggcgccgtgggtggagaatggcgccgcatatcgtgcagccggccgacactccgggcatggactc
tgaaccaccttgcaaattgggtatcctcatcaatttgctgctgaagcaaacactatcagaggagcactgccctgcaggtgcagagaa
caacttcagctgggatgggtctttga

>20668 OsMADS13

atggggaggggcaggattgagatcaagaggatcgagaacacgacaagccgaggtgaccttctgcaagcgccgcaacgggcttctca
agaaggcgtatgagctcctcctcctgcgatgccgaggtggctctcatcgtcttccagccgtggccgctctacgagtactccaacaac
aacaatgtgaaggctacaattgacaggtacaagaaggcgcatgcttggctcaacttctggtgacctctcatagaggtcaatgctcagca
atactaccagcaggagctcgccaaactcgccaccagaltcagatgctgcaaaacaccaacaagcactgggtggcgataatgtgagcaa
cctgtcactgaaggagctgaagcaacttgaagccgcttgagaaaggcattgcaagatcagagccaggagaatgaactgctggcttc
agagatcaattacatggccaaaaggagattgagcttcagaacgacaacatggacctcagaaccaagattgctgaggaggagcagcagc
tgcagcaggtgacggtggcccggtcgccgcatggagctgcaggtcgcgggcgggcgagcagcagcagcagaatcggttcgcg
gtggcgggcgcgagttggacatgaagtcttctccggtgaacctgttcgagggcgggcgaggtgcagggcggtggcgggcgagcgc
ccagcagatcatccaccgagctcaacctcggtaccaccaccacctcgccattccggcgccaccgcccgcgacgcgcccctcctc
acttctga

>20698 OsFDRMADS8

atggtgcgggggaggacggagctgaagcggattgagaacccgacgagccggcaggtgaccttctcaagcgccggaatggcctcctca
agaaggcgttcgagctcctcctcctgcgacgccgaggtgcctctcatcgtcttcccccgcgccgctctacgagttgccagcgc
ccccagcctacagaaaaccatcgaccgctataaagcatacacaaggatcatgtcaacaataagacaattcaacaagatatccagcaagtc
aaagatgatactttaggttggccaagaacttgaagctcttgatgagtcagacggaaaatattgggagaaaattagaaggatgctctattg
aagaactgcgtggtctagaaatgaaacttgagaagagcctccacaacataagactaaagaagaccgagcttctggagcggcagatagcca
agctgaaagagaaggagcggactttgcttaagacaacgaaaatttacgcggaaagcatcgcaaccttgaggctgcggcgctggtggcta
accacatgacgacgacgacggcgccggcggtggcgccgggacgtgcctatgacgagcagcagccggcgccgacgcccagc
gacgtggagactgatctgtacattggattgcccggcactgagcgtcctccaaccggctcgagacaggttga

>20700 OsMADS3

atggggagggggaagatcgagataaagcggatcgagaacacgacgaaccggcaggtgaccttctgcaagcgccgcaatggcctcctga
agaaggcgtacgagctgtcctcctcctgcgacgccgaggtgacctcatcgtcttccagccgcgggcgctctacgagtagccaacaa
cagtgtaaatccaccgttgagaggtacaagaaggcaaacagtgaacctccaactctggcacagttgcagaagtcaatgccagcacta
ccagcaggagtcctccaaactcgcccaacaatcagtagcttacagaacgcaaacagtaggaccatagtgggggattctatcaacacat
gagcctcagggaccttaaacaggtagagaacaggctggagaaaggcatagctaagataagggtagaaagaatgagctgttatatgctga
agttgagtacatgcagaaaagggaagttgagctgcagaatgacaacatgtacctgaggagcaaggtgttgagaatgagaggggacagc
agccactgaacatgatgggggcagcatcaacaagtgaatacgtatcatgttaataaccatagattccaggaacttcttcaagtgaacat
catgcagcagcctcagcattacgccatcagctgcagccaactaccttcaactcgggcagcagccggccttcaattag

>20770 OsMADS5

atggggcaggggaaagtagagctgaagcggatcgagaacaagataagccggcaggtgacgttcgcgaagaggaggaacgggctgct
gaagaaggcgtacgagctgtcctgctcctgcgacgccgaggtgacctcatcgtcttccaccgcgccgctctcagttctccacct
cctcctgtatgtacaagacactggagcgataccgagttgcaactacaaccttaactcatgtgaagcatctgctgactggaaactgaactaa
gcaattaccaagagtacttaagttaaagacaagagttgagttcctacaaactcagagaaatcttctggcgaggacttggtccacttag
cttgaaggagctcgagcaacttgagaaccagatcgagatatccctcatgaatatcaggtcatcaagaatcaacagttgcttgatcaagtatt
gagctcaaacgtaagggaacaacaactcaagatgctaataagacttaaaaaggagatacaagaaactagtgagagaaaatgcttcatat
atcttccaagatgtagggcccagtgccatgctagtgaagctaaccaagagtttccatcatgcaattgtgaccttccctgcatataggg
tatcaagcttacatggatcacctcaaccaatga

>20778 OsMADS8

Figure 10
Page 2 of 6

atggggagagggagggtggagctgaagaggatcgagaacaagatcaacaggcaggtgacgttcgcgaagcggaggaatgggctgctc
aagaaggcgtacgagctctccgtgctctgcgacgccgaggtcgccctcatcatcttccaaccgcggcaagctctacgagttctgcagcg
gccaaagcatgaccagaactttgaaagataccaaaaattcagttatggtggccagatactgcaatacagaacaaggaaaatgagttagt
gcaaagcagccgcaatgagtacctaactgaaggcacgggtggaaaattacagaggacccaaaggaaatcttcttggtgaagatcttg
gacattggcataaaagagctagagcagcttgagaacaacttgattcatccttgaggcacattagatccacaaggacacagcatatgcttg
atcagctcactgatctccagaggagggaacaaatgttgtgaagcaataagtgccctcagaagaaaactggaggagagcaaccagttgc
atggacaagtgtgggagcacggcgccaccctactcggctacgagcggcagtcgacctcatgccgtccagcaggtgccaccgcacgggtgg
caacggattcttccattccctggaagctgccgccgagcccaccttgagatcgggtttactccagagcagatgaacaactcatgcgtgactg
ccttcatgccgacatggctaccctga

>20837 OsBAA81880

atggcgcgaggagagcgaggagataaagaggatagagcgcggcgcgaggcaggtcaccttctccaagcgccggcgcgccctgtt
caagaaggccgaggagctctccgtgctctgcgacgccgacgtcgctcatcgtcttctccacggggaagctatcccacttcgaagc
tccagtatgaatgaaatcatcgacaagtacaacacattctaataatctgggaaagcagaacagccttcgctcacttgattagaacata
gcaagtatgcacattgaatgagcaacttgcggaagctagtctccggcttagacaaatgagaggtgaggagcttgagggttgagcattgat
gaactccagcagctagagaagaacctggaagctggtctgcacagggtgatgctgacaaaggatcaacaatttatgagcagatcagtgaa
ctccagcggaagagttcacagctggcagaggagaacatgcaactcagaacccaggtatccagatatcaccagctgagaagcaagttgtt
gatactgaaaattttgtaccgaaggacagtcctctgaatctgtgatgactgcattctggaagttcacagtcgcaggataatgatgatgg
ctcggatgtatccctgaaattagggtgccttggtgcatggaagtaa

>20842 OsMADS15

atggggcgggggaaggtgcagctgaagcggatagagaacaagatcaacaggcaggtgacgttctccaagaggaggaatggattgctga
agaaggcgacgagatctccgtcctctgcgacgccgaggtcgccgccatcgcttctccccaaagggaagctctacgagtacgccactg
actccaggtatggacaaaatccttgaacgttatgagcgtattcatatgctgaaaaggctcttatttcagctgaatccgagagttagggaaattg
gtgccatgaatacaggaaacttaaggcaagattgagaccatacaaaaatgtcacaaacacctcatgggagaggatcatgaatccctgaat
ctcaaagaactccaacagctagagcagcagctggagagttcattgaagcacataatatcaagaaagagccaccttatgcttgatccattcc
gagctgcagaaaaaggagaggtcactgcaggaggagaacaaggctctgcagaaggaactggtggagaggcagaagaatgtgagggg
ccagcagcaagttagggcagtgaggacaaacccaggtccaggcccaggcccaagcccaagcccaagcccaagccaagctcctcctc
ctccatgctgagggatcagcagcacttctccaccacaaaatactgctacccgccggtgatgatggcgagagaaatgatgcggcgcc
ggcgggcgcggtggcgggcgaggggccaggtgcaactccgcacggaggtcttccgccatggatgctgagccacctcaatgc
ttaa

>20847 Os008339

gagggggaggggagggcgggcgacggggaagagggagaggatagcaatacggaggatcgacaacctggcgggcgaggcaggtgacc
ttctgaagcggaggagggggctgttcaagaaggccgaggagcttccatcctctgcgacgccgaggtcgggctcgtcgtcttctccgcc
accggcaagctcttccaattcggcagcaccagcatggaacagattatgaccggtacaactcgactccaagacacttcagagagcagaac
cttctcaactcgacttacaaggggagggacagcagctactgtgccagactaaaggaggagcttgcagaaactagccttaggctgaggcagat
gagaggagaggagctgcacaggctgaatgtggaacagctgcaggagctagagaagagcctcgagtcgggtctaggtctgttctcaaga
ccaagagcaagaaaattctggatgagatcgatggactggaacgaaaggatgcaattgatagaggagaatttaagggtgaaggagcaa
gtgtccaggtatgcaagaatggaggagatgcagcctgggctgattcagaatcgtgtacgaggaagggcagtcgtctgaatctgtcacca
atgcttcttaccgcgacctccaccgacaacgactacagctctgatacatcttcaagctcgggtgcactcttaa

>19766 OsFDRMADS6

atggggcggggcaaggtgcagctgaagcggatcgagaacacgatcaaccggcaggtgaccttctccaagcgaggtcggggctgctca
agaaggcgaatgagatctccgtgctctgcgacgccgaggtcgctcatccttctccaccaaggggcaagctctacaagtacgccaccg
actcatgtatggacaaaatccttgaacgttatgagcgtactcctatgcagaaaaggctcttatttcagctgaatctgacactcaggggcaactg
gtgccacgaatataggaaactgaaggctaaaggttgagacaatacagaaatgtcaaaagcacctcatgggagaggatcttgaatcttgaatc
tcaaagagctgcagcagctggagcagcagctggaaaattcgttgaaacatatcagatccagaaagagccaactaatgctcgagtcattaa
cgagcttaacggaaggaaagtcactgcaggaggagaataaggtcctacagaagaactgggtggagaagcagaaagtcagaaagcaa
caagtgaatgggaccagacacaacctcaacaagttctcatcctccttcatgatgaggaagcccttcaacaactaatatcagtaac

Figure 10
Page 3 of 6

taccctgcagcagctggcgaaaggatagaggatgtagcagcagggcagccacagcatgttcgattgggctgccacatggatgctgag
ccacatcaacggctaa

>19877 OsRP5

atgaagatcattttcgtatttgcctccttgcattgttgcacgcttctgcacggttgatgctcttagtcaaagttatagacaatatcaacta
caatcgcatctcctgctacagcaacaagtgtcagcccatgcagtgagttcgtgaaggcaacagcatagtggaaccccccttctggca
accagctacgtttcaattgataaacaaccaagtcatgcagcaacagtggtgcaacagctcaggctggtagcgcaacaatctcactaccagg
ccattagtagcgttcaggcgattgtgcagcaactacagctgcagcaggtcggtgtgtctactttgatcagactcaagctcaagctcaagcttt
gctggccttaaacctgccatccatgtggtatctatcctaactactacattgtctccgaggagcattcccaccgttggtggtgctgtgactga
>20910 OsMADS14

atggggcggggcaaggtgcagctgaagcggatcgagaacaagatcaaccggcaggtgaccttctccaagcgcaggtcaaaactgctca
agaaggcgaatgagatctccgtgctctgcgacgccgaggtcgcgctcatcttctccaccaaggcgcaagctctacgagtacgccaccg
actcatgtatggacaaaaccttgaacgttatgagcgtactcctatgcagaaaagtccttatttcagctgaatctgacactcaggggcaactg
gtgccacgaatatagaaaactgaaggctaagggttgagacaatacagaaatgtcaaaagcacctcatgggagaggatcttgaatctttgaatc
tcaaagagctgcagcagctggagcagcagctggaaaattcgttgaaacatcagatccagaaagagccaactaatgctcagtgctcattaa
cgagcttcaacggaaggaaaagtcactgcaggaggagaataaggctctacagaaagaactggtggagaagcagaaaagtcagaagcaa
caagtgaatgggaccagacacaacctcaacaagtctcctcatcctcctcatgatgagggaagcccttccaacaactaatatcagtaac
taccctgcagcagctggcgaaaggatagaggatgtagcagcagggcagccacagcatgaacgcattgggctgccacatggatgctga
gccacatcaacggctaa

>20912 OsMADS18

atggggagagggcggtgcagctgcggcggatcgagaacaagataaacaggcaggtgaccttctccaagcggaggaaacgggctgctg
aagaaggcgcagagatctccgtgctctgtgacgccgacgtcgcgctcatcttctccaccaaggcgcaagctctacgagtctccagcc
actccagtatggaaaggatccttgaacgttaccagcgttactcgttgatgaaagagccgtactggagccaaatactgaggaccaggaaaa
ctgggggtgatgaatatggaattttgaagtccaaactggatgcacttcagaagagccaaaggcaactcttaggtgaacaattggacacactaa
caacaaaagaactccagcaattggaacatcaactggaatattcttgaagcatataagatcaaaaaagaatcagcttctgttgaatcaatttct
gagcttcagaagaaggaaaagtcacttaaaaaccagaataatgttctgcaaaagctcatggagacagaaaaggagaaaaacaatgctataa
taaacactaaccgggaggagcaaaatggagcaacaccaagcatcatccaacaccagtgcaggctccagatcccacccgacaaca
aataacagtcaaaaccaacagaggatcaggggagtcagaagctcaaccgtctccggcacaagcaggcaacagcaagcttccgccat
ggatgctccggacaagtcacacatga

>20914 OsMADS17

atggggaggggaaggggtgagctgaagcgcacgcagacaagatcaaccggcaggtcaccttctccaagcggcgcaacggcctcctca
agaaggcgtacgagctctccgtgctctgcgacgcggaggtggcgctcatcttctccagccggcgcaagctgtacgaattcggtagcg
ccggaataaacaagacattgaaaaagtacaatagttgctgttacaacgctcaagggtcaaatagtgctcttctggtggtgaacatcagagct
ggtaccaagagatgtcaaggctcaagactaagcttgaatgtctcaacgctctcagaggcacatgcttggtagaatcttgaccattgagc
ataaaggaaactgcagcagctggagaagcaacttgagtactactgtcacaggctcgacaacgaaagacacaaatcatgatggagcaggtc
gacgatcttcgccggaaggaaacgccagcttgagagctcaataagcaactgaaaaacaagctagaagctgaagccgatagcagcaactg
cagatcagccatccaggattcctgggtccatggcaccgtcgtcagtgccggcagagtgctgaatgctcaaccaccaccagatattgactgt
gagcctactctgcaaatgggtactatcaatttgcctgctcctgaggcgccaatccaagaagcaatggaggaggaggggatcagaacaac
aactttgtgatgggatggccctctga

>21116 OsMADS7

ttggttggtggcagctggagatcgatcgggatggggagggggcggggtggagctgaagaggatcgagaacaagatcaaccggaaggtg
acgttcgcaagaggaggaatggcctgctcaagaaggcgtagcgtctccgtcctctgcgacgccgaggtcgccctcatcttctcca
accggcgcaagctctacgagtctgcagcaccagagcatgactaaaacgcttgagaagtatcagaaatgcagttacgcaggacccgaaa
cagctgtccaaaatagagaaagtgaagcaattgaaagctagccgcaatgaatacctcaaaactgaaggcaaggggtgaaaatttacaacggac
tcaaagaaatttctgggtccagatcttgattcattaggcataaaagagctcgagagcctagagaagcagcttgattcatccctgaagcacgt
cagaactacaaggacaaaacatcttggtcgaaccaactgacggagcttcagagaaggaacaaatggtttctgaagcaaatagatgccttagg

Figure 10
Page 4 of 6

agaaaactggaggaaagcaaccatgttcgcgggcagcaagtgtgggagcagggctgcaacttaattggctatgaacgtcagcctgaagtgcagcagcctcttcacggcggaatgggtcttccatccactgtatgctgctgtgaaccaccctcagattgggtaccctgcagagcatcatgaggcgatgaacagtgcgtgcatgaacacctacatgccccatggctacatga

>22834 OsPN22834

cacgcgtccgatggctgaacacagctagctagctgggctaggatcgccatggatcggggtgaccaccaccacctccagcagcagcacca gttcttgatgccaccgcctgcgccgggtggtgccgccgagctgtgcatgccggcgatgatggcgagcagcagtagcatggacctggggcg gtggcgggggcgggcgggcgccggggagggcgggggccggggagaggaaagcgggcggttcacggaggagcagatacggctcgctg gtagtccatgttccacgcgcaccacccaagctggagccccgcgagaaggcgagctggcgcgagctgggcctgcagccgcggca ggtggccatctggttcagaacaagcgcgcgcggtggcgctcaagcagctcagcagcactacgccactccgctccaagtacgac gccctccactccgcgtcgagtcctcaagcagaagaagctcgtctcaccgtccagttgcacgagctaagagagaggctgagggaacg ggaggagaggagcggaacggtggtgccgccacgacggcagcaagcagcagcagctgcaacggatcaggcagcgaggaggtcgac gacgacgacgacaagaggaaacgccgccgggggtgctggacctgagccgccggagagctgctgctggcgggcgacgtgcg ccacgccggcgagcgtgctggtggagtcgagtcagtcgacgaccagcttgactacgatgagggtgtgtccggagtccttctgcgcca cgccggagctgtgggagccatggccgctcgtcagtggaatgcgggtggtga

>23495 OsPN23495

atgagctccggcgccgctcgcggcgccgccggcgccggcgccggagggaggggtcacctcgacgtggatcccttctc gtcgaggcgctcgacaacctcgcaccgcctcatggtttgcgtatggcaatggatatacagaatttatgcagaatcctcaactgcatgaa ttcgaattccagcatttccaacttctatctccgctgtgctgctcatcgtgttcacaacactatggttagagactacagtagcagatagcttag tagatggtcagtttagcaagatagttgcaaagaaaacatctgaaagcaaactccagttattgcatgtcagaagttccagcaaaacagcaa gaaatgagcatgaagctgcagagaagcttaagttgttattgtccaaggcccaaggcttccaaaatggtgcaggtgatccgggtgccaag aacaatgcagcaagaactgtgaagagaggatagaggaaatacaacaagcagcagcagcattttcaacggttccatttctgatattgaagg cacaagtgtttgggagcttattctgttcgagagatgaaccaattaatgttgagcctcctgttgatgagaataagggtgaacacaatgaacagc cgttcagagttgctgttttaaggacactgaaaaggatcgtagtgatcctgactatgatcgtactacaagaggtatgttaggggccagtgac acgacttcaatgtgagtcctggtggcttcaattttgttgccacagtttatgcagtatggtgtgttttatgcagtcgtctaatatgtcaagaaac caaccttctgtgtactttggccaacccgatttatctatgggctccttcttgaactgctgtctaccacagtgcccaactccagcaatgataac cccattgctatgacaacttgggcatatgatttctcaggttccgggttaccagtccttcaacctatggctga

>28517 OsBAB56078

atgggttggcggttggcagcagcagggcgatgacggcgggcgccggtggcgacatccccgacctcgccggcgggcgggcgaggcg gtagtggggagcgctgcgccacgcgggggtggtgcagctcgggtgccacacggaggaggtggagcccgccgcttcgtccgcaagtgc gagaagaccgagcagctcctccgcgactgcgtcgagcagtagacatagaagctcttgagaaaggcctttcgggagcattggttagctttctg gatgatgctgagaggatgaccaatgatttctgaagtctttgtgtcccttccatcaatgaaaggagtcgagctcatttgatggacaacctac aggcaggcacattggtggacaacctgcaggcaggcacattgaggaaggtagtgcgaaggacactaaacagaacgactacgcagaattca gcagcaagattacagatgtgtaa

>29949 OsPN29949

atggggagggggaagatagtgataaggaggatagacaactcagcagcagcaggtgacgttctcgaagcgtcggaaacgggcttctga agaaggcgaaggagctatccatcctctgcgatgcggagggtcgcccttgcgtcttccagcaccggcassctctatgatttccagcacc aacatgaaaactgtgatagaccggtataccaacgcaaaggaggagctacttggcgggaatgcaacttcagaaattaagatttggcagagg gaggcagcaagcttgaggcagcaactgcacaacttgaagaaagccacaagcaactgatgggtgaggagcttctggcctaggtgttaga gacctacaaggtttagagaataggcttgaataagctacgtaatacagaatgagaaggacaatctttgaaaagtgaatcgaggagtta catgtgaagggaagcctaattcaccaggaaaacatcgaacttctagaagcctaataatgtcatgtcgcaacaaaattggaactgtataacaag cttcaggcctgtgaacagagaggtgccacagatgcaaatgaaagtccagcactccatacagcttctgtatcatacaaaaatgctaataatgcct cctagtcttgaattgagccaatcacagcaaaagagaaggggagtgacgcaaaacagctgctccagaactgggacttcatctgccttaa

>29971 OsPN29971

gagaagagtgatgtgcaagtaccttggaatggagcttgatagaagggtcaaatgattggtcagtcgaactggcagaattccaatcagaaga acagcgattacgggaaagagtaagagagcttgacagcagaatgtatcttttcaagagaagttactttacttgaatcgaacagaattgatgtt

Figure 10
Page 5 of 6

tctaataagataacaagtttagaactgcaaaacaacagctcaatgacgagctacagaaagtaaagaagagcatgatactctttgaagtca
tcggttgagttgaatgataatttaacgaaaactgcagaggggtatg

>31165 OsPN31165

atggccgccgccgccgccgctcatcaccgccctccactccttcccgtggtctcctccgcagccggcggggatggccacctg
agcctttcgctcctcgcgcccccggcgccgggctgcagggcctcggcgccgacgttcaggggggacccgccgacgtacgcc
cgcgagatggagcgctcctccgccaaggagtcctcctcctcgccttcagagatgctggagggttgatccttggttagtgaaagactac
agggatgcagaagattgatgcaatgagcggattgtgggctcgagcgtctaaacccactccacggcccacaacatccttctggaag
gtcgtatggaacttgatgtgttggtgacagcagtcctggagcacttcagcccgtcttctgttgagaggtcacctacaactgtgcacactt
acaggactgatgtcctgatcaaggatggatactccaaaatttctccaacgtgaagttctaaatacgggtgcaaagcaagttcctgtgactac
tcagttgtctgtggagggtcctattaggtgaaagagggaatatgtcgagggcctaatagagatccctaggatcagagaagaacactgcctg
atcaactaaagggttcttggacaaactgcaggtgctctgcaacaactccctgccccataaggggacgtgttcagaggggattaaattgc
cacttaattgggatgttcagcgcctattcatgatttctacttgatgaggaaatactgattatcagagatgcatctggagcacctgatgtcctaa
caagattggaagggccacagccaaattcaattgatggcacatcagacgcagtgctgtcagaatatgaaagctag

>21044 Os018989-4003

atggcgctccttgcggcgatgccgcggcggtgcctccgccgcgccggcctggccaacctgctcattcgcgaaggcgccggcctcc
cttcgcgtcccgaacggtagcccgccattccggccctgtacttctgattccttctcctaattcttagggaaggggacgatattccccccaaaa
aaaatcggttctttacgcagcggcgggcgggcaatgcggcagagagggaggaggcggtgcgaacaggaacgggaagaaggaga
agaccggggcgagaggatcaccgggtgggggtcctgtgagttcagcaagatagtttctaagaaagttagggccaaaggaagaaccac
atataatgaggttccgatgagattttgcggagctgaagtcattacgcagaacgggtctggagttgatgagaagaatattaggcggagggt
atatgatgcttcaatgtgctcattgcaattcgtgttattgcaaaagataaaaaggagataaagtggatgggccttactaattatagatacga
gatacagaagttggaggaagttcacaagaactcatcaccaggatcaagaataagaagaagccttccaggaaattgaaaagcagtttgat
gaccttcagaatattacattacgaaccaggctagtcagaggccagcagaaggttgaatggcatcctccttccgttcttattgatcaagacat
ccgaaaagcaagggtggaaattgagatttcggaagattcgaagttgcacggctcgacttcaacgggtgcaccattcaccatgcatgatgat
gtatcaatccttgaagccatcaggcgtaacaaagggaagagctggcctctccattacccttaa

Figure 10
Page 6 of 6

Figure 11 (from HOS59 disclosure Example IV)

>19697 OsTFX1

atgagtcatagttcggatgaagattcggaaatcagtgattctgagatcgacgagatgcagacaaatttatgcacgactgggtggcagggga
 attcaaagtaaaggatgggcagagttacagctgccattttgagtggaagaagaagaagatttcaacataaacaatttattcagcatgc
 ctgaggagtaggcgctgcgtctaatcggcaggctaaagataaggcaacccatcgtgcccttgccaagcacttgaagaatggcctcactaag
 tcgtcaggacaacagtcacagacagctgtgtggagccacaaccgcttccaaacagagatgagaagttcgtgtggccctggatgggtgttc
 tagttaatgtgcctaccgaatggaaggatgggcgccaattggacgaagtggaaatcatttgaaggagcaactgtcacgctttgccactta
 agatcattccattatggaattttagaggtcattcagggaatgctattgttgagtttgaaaaagattggcatggcttcagaaatgcgcttgccttga
 agattactttgggaaagaaggatgtagtaggggactggaaggagaaacagaatcaagggtcaaacctcttggatgggttgcagagct
 gaagatcacacttctccagggttaataggcgaccatttaagaaaaaatgggtgacctgaagactatcaatgatcttgagaatgaaggagcacg
 taaaactgacaaactttagctaattagcaaaccaattgaggtaaagaacaggcatttgcaggaacttgaagttacatacaatgaagaac
 tacatcacttgagaagatgatggggcaagggaacagcttctccagaaataatgaagaaattcgaagatgcagcagctagctcagaga
 cattctcagaagataatcgtatgagaatcaaaagctgcgttctgaactggagtgcaagatgagtgaacttaatacagatccaaggagcttga
 gagattgctgcaaaaagtattatgaccgaaggatcattgatcaagagaagcagaagaatgctatcaaatcaagccatttaagttggcaac
 gttggaacaagagagagctgatgaaaatgtgctgaagcttgaagagaacataagagagagaaagaagctgctgtaaagaaaatttgaag
 ttagaacagcaggttgatgcaaaacaaaagcttgaattagatatacagcagttaaagggcaaatggaagtatgaagcatatgccagggg
 atgaagattcagcattgaaaaataaaattgatgaattgagtgaggagctgcaagaaaagatggatgaactagatgctatggagtcacttaacc
 aaactttggttattaaagaaagaaaaagcaactactgaaatgcaagatgccagggaaggagcttgaatggcttacttgaatctttagatggtca
 atcacatataaggaataaagagaatgggcgagcttgatctggaagcattttcaaaagcatgcagaagatgtcatcggaaggagatgcagag
 attactgctgccattctttgttcaaaagtggcaagctgaaattaaaaatcctgattggcaccttttaggttgtcttggatgggcaggagaag
 gaaattatcgaggatgatgcaaaacttcaagaattaaaagaagagcatggcgaagacatacagattggtgagggacgcgctgtgtgaaa
 ttaatgagtacaaccttagcggcagattcccagtaggagagctgtggaacttcaaggataaacggaaggccacttgaaggaaacgggtcc
 agttcgtcctgagacagtggcgagcgaacaggaggaagcgctaa

>20080 Os005792-3529

atgggtcgagctcctccagctcgccatcgactgctcggcgagcaccggcagagcgccggtccatgtccgaggtggccgcgaggatcga
 cgagatccgccgctccagcctcggcgaccgtccggccaccgacagcgccggcgaggggcaggagccctcactataa

>20257 OsCYCOS2

atggagaacatgagatctgagaactttaaccaaggtgtttcaatggagggcgtaagcatgcgcccagatggcaaacaccaacaggagg
 gctctcggggacatcaagaacatcatcggagcacctcaccagcacatggctgtgagcaagaggggtttagataaaacctgccgctaag
 aaccaatctggacaccgtctatgacgaggaatttctgcaacactggcaaaccaacctccattgctcccctggcaccattggaagtga
 gaggcaaaaaaggacagcagattctgatttcatgggtccagcagatatggagtgcaccaagatcacatctgatatttgcattgccaatgat
 gtctgagatggacgaagtgatgggttctgaactgaaagaaattgagatggaagacattgaggaggcagcacctgatattgacagctgtgat
 gcaataaactcccttgcagtagttgaatatgttgatgaaatttacagcttctacaggaggagtgggtttagctgtgtctctcctaattacatg
 ttgagccaaaatgacataaatgagaagatgcgcggcattctatcgattggctgatagaggtgcattacaagttagagctattggatgagacc
 ctcttcttactgtgaatatcatagaccgattcttggctcgcgaaaatgtggtgcgaaagaagcttcagttggttggcgtgactgccatgtgct
 cgcctgcaagtatgaagaagtgagcgttctgtatggaggatcttaccataatctgcgaccgcgctacacaagaacagatattctcgaat
 ggagaggatgattgtaaacactcttcagttgatatgcagttccaactccatactgtttcatgagaaggttctcaaggctgcacaatctgaca
 agaagcttgagctcatgtcttcttcataattgagctgagccttctcagatgatgagatgctcaaatccagccgtaatgctggcggtgctgcc
 atctacactgctcagtgaccattaatgggttcaagtcctggaacaaatgctgtgagctgcataaaaatactctgaagaacagctgatggag
 tgcaccaagatgatggttgagctccacaaaaagcaggacatgggaagcttactggagttcatagaagtagacacatttaggtatggttg
 ccccgcaaaatcgagggccggtcttcttctgctcaagagcgtggcactgtaa

>20466 Os005750-3115

atggcagatgctagttcgaggactgacacatcaacagtcctggataccgacgataagaatcagatggtagacgggcaagtgagctattg
 tgccttcaattcatctgatcggtctgacagatctgacagcccatggacaaaaaggtgttacggcggttctcaaaatctgaggctgcaa
 gaaaaagtcggctgagaaaaaggcatatgtacaacaattagagagcagtaagctgaaacttgaagcttggagcaagagatcaataag
 ctgccacaaggaatttcatctcagagctcaggagaccaaacatcatgctatgagtggaatggagctatgacttttgattagaatatgccccg

Figure 11

Page 1 of 7

ttggttgagggaacaaaacagcagataaatgagctgaggactgcagtaaatgctcatgcaagtgcagcgacctccgtctcattgtagatg
 ggataatggcgacattacgatgagatattcaggctgaagggtgttgcgcgcaaaggctgatgtttcatatactttcaggcatgtggaaaacac
 ctgctgaaagggtgcttctgtggcttgggggttttcgttctctgagcttctaaagcttctgtgaatcagctcgcagccattaaactgagcagcagt
 tgttgggactatcgaacctccaacagtcctctcagcaggctgaagatgctctatcacagggaatggaagcgttgcaacaatccttggcagat
 acgttggccgggtcccttggccatcaggatcttcagggaacgtggcaactacatgggtcaaattggctatggccatgggcaaaacttggga
 cccttgagaatttctccgtcaggtgacaattgcggcagcagactttacataaatgcagcgaatttgcacaatccgacaagctgcccggtg
 ctctacttgaatacatgattacttctcacgttgcgtgccctgagttctcttggcttgcaggccacgggagtaa
 >20534 Os018049-3655

atgcagcagcagcagcagcacctccaccaccagccgtcgcgcggcgggcgcggtcgcagctggcagctgcagcagggtgcgtg
 gcatctccgcggctcaagggttccctcccgacgcagcagctcctgcaggagtctgcagcctccccgtgaagagcaccaccagcccatc
 atcggaagcaaggcaaccaaacccgcgcaagaaggccgcccagcggcgagggtcgtcgtcgtggaccgcggcgacgcagatcc
 agtccatggacgcggcgagctgcagcggctcaagggtcaagctctacacatgctcgaagggtggataggaggtacaggaggtactg
 cgagcagatgagggcgctcgcggcgtcgttgcaggcgggtggcggggagaggcgggcggtcgtacacgcgggtggcgtcagga
 cgaatcgcaggcacttccggagcctgagggacgggggtggcgcagctgcaggcgggtgcggaagcagctcggggagaaggacacg
 gcgggtgcccgggatgaccaaggcgagacgccccgcctccgggtgctcgaccagtgcctccgccagcacaaggcctaccaggccggc
 atgctcgagagccaccatggcgccccagcgcggcctcccgagcgcggcgtctccatcctccgtgcctggctcttcgagcacttcttc
 acccgtaaccaagcgacgtggataagcacatcctcgcgaggcagaccgggtatcacgtagtcaattggcgaaltggttcatcaacgcga
 ggggtgagggtgtggaagccaatggtggaggagatgtacgcggaggagatgaaggacgaggagggcagtggaacagtcacgcagggcg
 agcaaccccgagcgcgtg
 >20559 OsHOS59

ctcgcaccttcatgaccattatgtattgctcctttgttctgaagaacaaactacagcaacatgtgcgtgttcatgcaatggaagcagtaatg
 gcttgcgtgggaacttgaacaaactttacagagccttacaggggcatctcctgtgaagggttctggagcaactatgtctgatgacgaagacaat
 cagggtgatagtgcagcaacatgtttgatggaatgatggatcagatggatgggttggcccttaattgtcagggaggcgagagatc
 attagtgcagctgtacggcaagagctgaaacatgagcttaaacagggggtacagagaaaagcttggacattagggaagagataactccga
 aagcgaagagctggaaaactcccaggagatacagcgtctacttgaagcatggtggcagggtcactctaaatggccataaccaactgag
 gaggacaaggctcgttgggtgcaggaaacagggttgaactaaaacagatcaataattggttatcaaccaacgtaaacggaactggcaca
 gcaatcctgcttcatcctcatcagacaaaagcaagagaaaaaggtatcgtgctgtggactttaa
 >20689 OsMYB

atggggaggcagccgtgctgcgacaagggtgggtgaagaagggtccatggacggcgaggaggaccagaagctcgtcagcttccctcc
 tcggcaacggccaatgctgctggcgtgccgtccccaaagctcgcgggttgcctcgggtgtgggaagagctgccgggtccgggtggacgaac
 tacctccggcggaacctgaagcgtggtctcctgtcggagacggaggagaagacgggtgatcgcacgtgcagcagctcggcaacaggt
 ggtccaagatcgcgtcgcacgtgccgggcaggacggacaacgatcaagaaccactggaacacgcacatcaagaagaagctccgga
 agatggggatcgacccgtcacacacaagccctctacctgctccgccgtggccgacggcggtcgcgggagcagaaggtgccgga
 ggaggaggaggaggtggaggagaagtcgtcgtcgtggaaagctcgcagcagcctgcgcgggccacgacgtgttctgcaccga
 cgagggtgccatgctgcacctggacgacatcgtcctccgcccggtgcgacgtcgtcggcgacaccgccgggtcgcgggcagagtcgt
 cgtccacgtcgcagctcgtcgcagcggcgggcgggcagcagcaggtggtcgtcggcgacaccgccgggtcgcgggcagagtcgt
 gatggggctcgcagcgtcgcacatggtcaccacggcgggcgccggcgatggcgacgtcgtgggagttcgaggatccgtcaacgcgtac
 cagaggatcgcgttgttgcacaccacagagctgacgtgggcttga
 >21036 Os003181-3684

atgtccacggaaagcggcatgctacgaggcgcaggcgtcgggtgggtctccggcgcggtcttctccatcagggccgtcgaatcgtgcac
 gaaatcggagatcaagtgaatcaggaaaatacagcattatctcgtgcttgacatcatcttagcctcttcagtgaaggattgtgtggaga
 aggtcagcccagcactgcagcgcgcgggtgcagagtcagatgagtcactgagtagacacattcattgacaacaatgaccttttgaaccgga
 aatacaggagcatgtaagagatctaattaacagaatcccaagaccacattcagcgtcgaaccaatcctgatcaggaaactgataactg
 ttgtgcagtttgccttcaggattttggagcatcgcaatttgcgggtctcgtcctcattgccagcacacgttccacgcacgatgcatcgacaact
 ggcttttcaggcagcagcatcatgcccgtatgcagggtggtgtgcatatagaccatatacatatgtaa
 >22896 OsAAD27557

Figure 11
 Page 2 of 7

atggattc gactaagcaagatttccaaccccgacattctccatcaaactgtggccaccaagt gaaagcactcgtctcatgcttgtggagagg
atgaccaagaaccttccacagagtcacatatttctcgcaagtacggccttttgggcaagggaaggaggtcatgataatgctaaaaggatcgaa
gaggtatgctttgctctgcagacgagcattcaaggaggagcctgatggggatgggagttctgctgtccagctgtacgcaaaggaaacta
gcaagttagtgctggaagtctgaaaagaggccaaggaccactgtggaaccagaagtacctgttgctgatacactcttgagcctgctgat
tctgtatttgatatactgggtggcaagcgtgcgtttattgaggcagatgaagcgaaggaaactttgagcccgttatcaaaccaggaaatgcat
ataaacggatttgcctcagcaacaggagccttgggtattggtgctgccaatgttgctggacattctggaatcaatcaaaaagcaactcacaga
ggtggatatacagatttggctggtgagacctgaggatgaagcccttgatgtgatgcgcactcttcaaaaagccttagaaggcgtgtactg
agatactgaacatctctgataatgcttaggtgagaagggtgtcagagcattgaagagctcctgaaatcacaggacaacctggaagaacta
tatgtgatgaatgatggtatatcagaggaagccgcacaagctcttctgaactattccttcaactgagaagccttaagatcctccacttccaca
caatatgacgggggatgaaggtgctatgtttattgctgagatggttaaacgttctcctaacttagagagttcaggtgctcagcaacaaggata
ggatctgatggtgggtggctgctgagcattagggacctgtactgctgaagaaacttgatctcagggacaactgtttggagttgag
gcagggttggctcagcaaaaccccttcaaaacttctgatctgttgagctgtatctcagtgatctcaatcttgagaacaagggcactgtagc
aattatcaataccctcaaacagtcagcaccccgattggaggctctgaaatggctgggaatgaaattaacgccaaagcatcccaagccttagc
agaatgcctgacagcaatgcaatcacttaagaagttgaccttggcagagaatgaactgaaggatgatggtgctgtggtgattgcaaaatcac
tggaagatggacaccaggatctgaaggagcttgatgttagcacgaatatgctgcagaggggtgggagctcgggtgcttggccaggcaatcg
caaacaagccaggtttgtgcaactgaacataaacgggaatttcatctcggacgaagggttgatgaggtgaaagataattcgaagagtggt
gagaactctgtggaggtgcttggccactagatgagaacgacctgagggggaggctgaagatgatgaggaagagggaagaggaggaag
aggatgacgatcaacggcgtggattgcgccaccggagcccgcacccacacccctcgcctcgaatcactctacaacagatcgga
acctcggggcgccggcgccggcgccggcggttaagcgagcggacaccggcgacgagcgcgaggaggaggaggaggaggaggga
gaagatggcgccgacgctgtcgcagctgcagcagggatcgtccgcggttgccatcgccggttaccgactacgctgggaag
ataaactgcctcacttccaccgcaaggaggatctcctcgtcacctccagcgaggacgactccatacgccctctacaacatcaccagccca
ctttacatgttctgttcccctggggcatcttgtcgttagacctcatattgtagatgttgacgttgatgcgagagtgggtccctttg
acacttcttgggttgggtgagatacagctgaagtttctgacatcaaatttagcaatgatggcaagtccatgctttgacaacaaccaaacat
atatatgttctggatgcatatggaggggataagaggtgtggcttagtctggagtcatctcccaatgtagcaactgaagctgctttactccaga
tgcccaatatgtaattcagattgcttgggaacagtcacatcggtcctatcactgcttgaagtgggtccccgtcgagcaatgtttgcaactg
cgtcaactgccctaacttctggattccaacaattccagttcgaattaggtgcctag

>23251 OsPN23251

atggggggccggaagcgcgctgctggtgggcatcaactaccggggcaccgaaggcgagctcaagggggtgccacaacgacgtggc
ccgcatgcgccgcccctgctgaccgcttggcttcgacgaggccgacatccgctcctcggcgacgccgaccggcggcgccgcag
cccacgggggccaacatccgcgggagctcgcgcgctcgtcggcgacgccgccccgggacttctcttcttccactacagcgcca
cgggacgcggtgcccggcgagaccggccaggacgacacccggctacgacgagtgatcgtgccctccgacatgaacctatcac
agaccaagatttcacagaactcgttcaaaaaggtccagatgactgcttaccatagatctgactcctgccacagcgggtggttacttgac
aagaccaaggagcagattgggcacagcaccaagcagaaccaggcccagcagatcaagcgagaagaacgttctgactctggcaccggc
ggtttccgactccttccaaagaaacccctcaaagaaacagtccgggacgcgttcgagtcgaggaattcatccctcaccagagcagcc
ggcgcaacgacgacgaggatgaggagccacacatgggtcctcctccatggtggcgatcgatcaagaaccgggtcattgcccctctcg
acgttgattgagatgctcaaggagaagaccggcaaggacgacattgacgttggctcgatccggatgacgctgttcagcctctcgggtgacg
acgcgagcccaagatcaagaagttcatgaaggtcatgctgaccaagctccaggaggccagcagcgccggtgatgggcttgggtggg
cgccctggcgaggagttcatgaaggctaagctggagggaaccaggagcgccgacgccctcgagccggccatgaacaggaggtgca
cagcgtccacgaggcgtacggcgacgagggcgaggtgagcaatggcgtgctgattagcgggtgccagaccgaccagacatcggc
ggacgccacgacgccaaggcggtgtcctacggcgctcagcaacgccatccagaccatcctgtcgagagaagagcggcagggtgac
gaacaaggagctcgtgtagagcgcgaggtgtgtccaagcaagggtacaccagcagcctggccttactgagtgacaagcacac
cagtggtggccttcatatgctga

>23253 OsAAK00972

atggctacttactcagaccctggcaatgaaagggactcgaagctatgtaccagcggattcagggaattcatcatatcctgtgccatca
gcaataggaaacatgtatatacctggcaatgggtcttctggccatacacggaattcagttggcattatccagcatcagcagaatttcatggag
ctgcttggccatccaactgcgatcttcaagattcatcgtcacgggaacctaacatggctcgctgtacatggatcagcgtcttttggacctg

Figure 11
Page 3 of 7

ccaaagatatgagaaatgagatgttgatgcatctgatggatggagcacataatgctgggtgatctcatccacaatgacactcatagcagcg
cgagattgagttggcctattgaacaaccacaattcgcggttgaccagcaccaggccaaggattgtcttgagcctcaacacgcat
atcctggcgcccttgatccatactggtctgcgaaaacagagttgctaaccacacactcttaccatggtgatgacaacagaatgaagaatatg
caatctgaggcctcacaggcaatcagaaactcaaagtatctgaaagcagcacaagaattgcttgatgaggtcgtcagtggttggaagagtat
aaagcagaaggctcagaaagaccaggctgaagcaggaaaatcagatacaaaagaagccgagggggggttcgaaaggtgaggggggtatc
ttcaaccacagaggatctactgccaatgctgcaccagagatttctgctgctgagaaacaagagctccagaataagatggcaaaacttatgg
ccatgttgatgaggtggaccgaaagtacaacactattatcatcaaatgcaaatgtagtctcatctttgatatggttgctgggtctggagctg
ccaagccttatactgcagtggccct

tcagacaatctcaaaacatttcagatgtctgaaagatgctatcaacgatcagatcaatgttatccggaagaaacttgagagaggaggagaggtc
atctggcaaaaggggaaaattaacgcgcctccgttatattgaccagcaattaagacaacagcgcttccagcagtatggttggttacagca
aaatgcttgaggccacagaggggactgccgaaaactcagttcaattctctgctgcttggtttgaacacttcttcacccgtatccaaaag
attcagaaaagctgatgctagcgagacaaactggcttaacaagaagtcagattcaaatggttcataaatgccgtgtccgcctgtggaac
cgatgatcgaagacatgtataaagaagagattggggaggcggatctcgactcgaactcctctccgacaacgtaccaaggagcaaggac
aaaatagcaacatctgaagataaggaagatctgaaaagctctatgagccagacttatcaaccaagccaattaggtgaatccaagccaacat
cgggatgatgagccttggtggggcaccagccggcttccacaacgaaggcaaccaggatgacagcttcatgaacctaatgtgaaggacc
aaaggccgggggaagccgaaggcagcctcctccatgatgccgtcgccatcattccgatgagaacgctcggttcatggttaccattgtc
ggggctcggagatacgggaacagcaatgigtcatgtgacacttggttacagatcctgacaacaggcttccggtacagaacactcatcag
ccaggtttcggcggtgctggagaagaaattacaactccaccgcttctcgggtgctgctgcagcttctctcggactacgaatccacgaac
caaatagatcaaaggcaacggttcgaaccatcgctctaatacatgattttgtggcctaa

>23388 OsPN23388

atggcggcgggcgaccggaggccgctgccgatggcccctcggcgattggtcatctgctgcaccgggctatccaaagagcgagg
actcaggtgaaggaggcgaaactgagaggatgggaggcgagtacagcgggagcctgcatcccaagtgcacccacctacgcgcagcaca
gcttcgccggccgaagttgagcacgcggtcaagcatggcggaagaatggcctcttgttgtaacacttggttggttggtgattgcgtgc
ggaggagcatgaggttgatgaatcgttatatagcatcaagaacattgggggagaatggaatgccactcggggagttcaaccggcttgcgg
cgctcctgttagtggaattcttgccctccaccgatgatgttcaggaaaaaacatttcagacacgaccgaaaagcatcggttcagacatct
agaaaagaacatgaccatgatgaatttctcttacaacgactcgtatcattgacctgggatacttggtgaaatgaggaagaagtatctga
atgctgctactagagaagggtgcaaaattgttgaccactggttattggctgctatgcgacatatgttggtgtaggagcgttctgtcaagaga
tatgttggtcattcagacaacattgaactccactttgatattgaaaaccgcaaaaggaaaagggttgacgcttcttccattgtctcggac
ctggctaggcaggttgccacgattcttgaatgcccagacattcaagagaatagaaaattggggatgttccttctgcaattcaaaactcca
gtggagtaccatcaaccaaggggagatagatgaattcatcaagagaggcaaaaattgttgaaatagcaaaagaaatgtccgagatcg
ccgtgctcgccgtatgcagtcagtggaagtacatccatccatcacgcctgtgaaacttatggaatcaatctgttgaccgtatctgagcca
actacgtccgcatgcatttatacggagtttcatgtcagacgatgcctttgagcaacaaagcactacttctcgtatgcgaatggggatggca
aggatgatcagtcagcagatagtttaccgccccactaagagaaagtgagaagagtgaagtatcttaagaaccacttcttaccgtactct
tccccattgatcgtttgtgagcttgacccttcaaggacattcttcagcaatggtggttcacacgcatacaagtgttgatcacatctaca
acttctaccaggaaaatatgtcttccgatgagataatgtagccctgcaaacggattcacgacatgctgatcagcttctcgtctgacgcaa
gtactgaatcagcagaaagaggtttgtaacattcaaaagaattgatttttgggaagccgaagaagtttgaggggctaaaacgcctcagca
gagagaacaatagcaatgtatatgagcttgattaggccataa

>23829 OsPN23829

atggcgctctccgtggagaagacctgctggggaggagtagaaggtgaaggacctctccaggcgaggacttcggccgctcagatcga
gctcggcgaggtcgagatcgccggggctcatggcgtgccgcgccgagttggccctccagccgtcaaggcgcccgatctccggg
tccctccacatgaccatccagaccgctcctcatcgagaccctaccgccccttggcgccgaggtccgctggtgctcctgcaacatctctc
cacgcaggaccacgcccgccgcccacatcgccagggactcggccgctgttcgctggaagggggagacctcagaggagtactggtg
gtgcaccgagcgtgctcactggggcgtcggcgccggcccgacctcatcgtcgacgacggcgccgacgccagctgctcatccac
gagggcgtcaaggccgagggaggttcgagaagtcaggcaaggtccccgaccggagtcaccgacaacgccgagttcaagatcgtg
ctcaccatcatccgcagcgccctcaagtcggacccagcaagtaccgcaagatgaaggagaggtcgtcgaggtctccgaggagacca
ccaccggtgcaagaggctctaccagatgcaggagaccggcgccctccttccccgccatcaacgtcaacgactccgtcaccaagagca

Figure 11
Page 4 of 7

agtttgacaacctgtatggttgcgcgacctctctccctgatggctcatgagggctaccgatgttatgatcgcctggcaaggttgccgtggctcg
cggttatggtgatgttggcaagggctgtgctgctgctcaagcaggctgggtgccgtgtcattgttactgagattgaccccatctgtgccctc
caggcccttatggagggctccaggtcctcaccttgaggatgttgtctcgagggtgacatcttttgaccaccactggcaacaaggacat
cataatggttgaccacatgaggaagatgaagaacaatgccatcgtttgcaacattggtcacittgacaatgagattgacatgctcggccttga
gacctaccctggtgtcaagcgcatcaccatcaagcctcagaccgaccgctgggtcttccttgagaccaacactggcatcattgtccttgctg
agggctgctcatgaaccttggtgctgctactggccacccagttttgcatgctcgtgctcaltcactaaccaggtcattgtcagcttgagctg
tggaaggagaagagcactggcaagtacgagaagaaggtgtacgttcttccaagcacctcgacgagaaggtggccgccctccacttggg
caagcttggtgccaggctgaccaagctctccaagtcgcaggctgactacatcagcgttccagttgaggggtccctacaagcccgcgactac
cgggtactag

>23830 OsPN23830

atgggggttcctcaggactttcaagccagtgttgaagcattgccagctatgctgcaaaggaactattcattaatgcgggagctagataaaagc
ttgcaaggtgtgcaaacgggaaatgagcagcgctgtcagcaagaaatagaagacattaaagcatggacttgaatctggaagtattacatatga
cccggcaaaagcttaaaattttctgatgaagcaatcgaggagcagaagcattgtgttcgaattgctgatgagaaggtggctttagccagtcagac
ttatgacctggttgatgctcatatccaacagctagatcagtttatgagaaaacttgaagaacttcgacaagaaaaggaggcagctacaactgc
tgctgctgctgctgctgcgggcggtgctagtgtgtgctacaggtacgcctgttgctgctactgtaacagctagcgctggtacttcaactgctgac
aatactccaaaagggtggaagggtctagcgaagaggccgaggggggtcgaagaagactgctaagggtgccactgagcaaccagcaccag
ccatagatttagaactgcctgtggatcctaacgaaccaacatattgcctctgcaaccaagtcagctacggagagatggtcgcatgcgacaat
aatgattgcaagatcgaatggtaacatttgggtgctgctggcgtgaaagagcatccgaagggggaagtgtgtattcccaagttgcattggattc
caaaagaagcgaaaaggaaagtga

>23832 OsBAB07943

atggcgaccttcgcgaacctgagaacgcgctcaagagagcggaggagtgtgatacatgttgggcagaagcaggcagcgctacaggcggt
acatgatctcatcacttcaaaaagatacaggtcatggcaaaagccactcgaaggattatgatgaagtattgtggagctctgtgtagacctgag
gaagggcaggttgcgaaaagatggccttattcagtcacaggattgtctgccagcaagtgaatgtgtcatccttggaggaggtgataaaacact
catgcagctttccaatgagaaagcagagcaagcaaggaaaccaggcgcaagcacttgaagatgccctggatgtagaagatctggaagca
gacaagcggcctgaggattgatgtcagctatgtcagcgggggagaaagggaaggaccgatcagacagagagcatgtcacaccatgggt
caagtttctctgggaacatacagaacagttcttgatattgagaacaattcaaagcttgaagcattatatgctatgactgtctacaaggcttt
tcaattttgtaagcagataagagaacgactgaattccgtaggttgtgcgagataataagaaatcatcttgctaactgaacaaataaccgtgacc
agcgagatcgtcctgatctgactgccccagaaagcttacagctatacctggacactagagttgaacaactaaaaattgctacagagctctctc
tctggcaggaggcctttaggtctgtggaagacatccatggcttgatgagcatgggtgaagaaaacacaaaaccatctgtcctgggtgtgtatt
atgccaaagtgtactgaaatattctggatactgattgccacttgatcatgcatatgcatggcttaagctatttacttgcagaaaagctacaataa
gaactgtctcagaaggaccttcaattaatagcttcttctgtttgcttgcacttgcctgtgcacatgataccataaatatggcgcatctcatc
ttgagcttgaanaatgagaagatcgtaatctgcggatagctaattctgtcaacttctcccttgactccaagcgtgagaatagagaagtgccatc
aagagcatctctttttctgagctggctgccaaggtgtcattgcctgtgcttcacaagatgtcaagatttatacaaccttttgaacacgatttt
cttctctggatctcgtatccaaagcacagccactgctatctaagatttcaaagattggagggaagcttcatcagctcccttgggtccagaagt
tttctgtcacagtaccttctgctttggagaagttaacaactttaagggtgttgcaacaggcttcccagatatttcagctctgtgaaaattgatatgc
tctcaagaatgatccattcttgacttttctgttgttgagaagatctctgttgatgccgtgaaacacaattttgtgtctatgaaagtcaaccatttat
caggagctgtccattttgggaaaaatggacattgaatctgattgtctcagtaatcacctaagtgttttggctgatttcttaataaagcaaggagc
cttatccatccacctgtgaagaagccatccaagcttgggtgaaaacttaacgagcttctgtgcagtagttggagaatgagcacaagcgtcttcttg
cgaggaaatcaattattgaaaaacgcaaagaagatcttgagcgccaaatatttgaaaaaggaaaagggaagaagaaaaaaagagattgagtg
ttctcaagaaatctgcagaagatgaaggataaggcttctcaatgatgtaaaattgagggaacaggagaggatccgtcgacaattgggtgaa
aaagaaaagatagaggcagaagaactgttcagaaacagataaaggagatagcaaaagagaggaggagaaaaaacctgtccttcagggggg
aagttactaaggagggtgtgatggaattggctatgaatgagcagtttaaggagcgccaagagatggagaaaaaactacagaagactggttaa
gcaaatggattacttgagaggggcaaaaaggcagggaagaggcgaccttcaattgagcaagcttttcaaaaacgccttgaagtagagaaaatc
cttcatgagcaagagcagctgagagagattgagctcagcaagcaacaccatgcagggtgacctgcaggagaagaacagactttctcgaatg
ttggagcataagaacattttccagggaagaattgtccaacgtcgtgaagctgagtttagtctgttaagaaaagagagggtgaaaggactagt
caattgataatcatcaagggaagcgtgaaagggatacagtgcggaagttaatgtactatcttaacttagaagagcagcgccctcaaaaggctccg

Figure 11
Page 5 of 7

ccagcaatattgtagattaccgaaaatgtgtcaacaggctatatctagtcagccgcatttagaccgcgtgccatgcaacctcagtgtag
 tccattagtcgaaaatgggcgtcctgtactacaaggcgttttggtgtcgtatataatagccaaagggtacttggatgggttcaattgatgccg
 aactatgccactgagcttgttagatgttcgcacatttttagaacgtacacatgaaagatgccgtgcatcctatatggaggctgtccttgaaaa
 acaaagttatattctctcaaggaaatgatgttgaaagttaatgcgcttggatccagcaaatctattacaaaattcttttggtcaacttgacc
 acagtattccagatgcagaagctgtagaagtagagattgaactaagtgtctcctattggatgtgcccataaaaacaggaaaacttgatcc
 atgatgaccagaaattgatactgttagcatctcttagtgattcattggaatatttagcggattcagttgaaaggcttggagagtcattcattaattc
 atctaccatgttggagaacaaaaatcacatccaccagggtcgcatactcgctcaactagtgaattccgaaaagccttgcatcacttgcta
 gtagtagaaggctagcaattgattgtgtcagagctttagattggaatgcaattagaactatatacacatgcaggaaatgacgaaaaga
 gaatatgtagaagatcaagatgtgagatcctgatgacttcatctcactaacaactcagatcgacgtaggatgaagaaatggctcct
 tatattgcagaatctaaaaggaaattatgtattgggtgggatttctagcgttgcgtgctaattgcatcgataaaggcactcgctcagatgaagtcgatc
 aatttggtaggggtccaacaaatagcagaattctatagcactagaacaggcttggcagccataccctcaattgacagtgaaagccgtccaa
 cagagaatagatcgtgttcggacattctacgaactctgaattgccttgcagcttcttggatttattgcagaacatgaatatttttctgca
 aaagaatacttgagcgtgttaaagggtcaacgttcctggagagagatgccaatggatgctgaacgacgcacatctcagatttgggccattaa
 >23169 Os000221-3976

ccacgcgtccggagagagagagagaagaaggatgctgaggaggggaaagggtgaggatgttgatgaggagaaggagaagaagagaa
 gaaaaagaagaagatcaaggaaagtctctcacgagtggaacgtgatgaacaagcagaaccatttgggtgaggaagccagaggagatca
 ccaagggaagagtacgctgcttctacaagagcttgacaaacgactgggaggaacatctggctgtcaagcacttctgtggagggtcagctt
 gaattcaaggccatcctgtttgtaccaaagagagcgcatttgaccttcttgacaccagggaagaagcaaaacaacatcaagctgtacgtacg
 ccgggtgtttatcatggacaactgtgaggagttgatcccagagtggtcagcttctcaagggcattgttgattctgaagaccttccctcaac
 atctcacgtgagatgctccagcagaacaagatcctgaagggtgatccgaagaaccttgtaagaagtgcgtggagctcttcttgatcgc
 tgagaacaagggaagactacaacaagtctacgaggcttctcaagaatctcaagcttggcatccatgaggattccaccaacaggaccaag
 attgctgagctcctgaggtatcactccaccaagagtggatgagttgaccagcctcaaggactacgtaacccggatgaaggaggggccag
 agtgagatctattacatcactggtgagagcaagaaggctgttgagaactcccccttctcgagaagctgaagaagaagggttatgaggtcct
 gtacatggttgatgccattgatgagtacgctgttggcagcttaaggagtttgaaggcaagaagctgtctctgccaccaagggaaggctgaa
 gcttgatgagagtgaggacgagaagaagcggcaggaggaactcaaggagaagttcgagggtctgtgcaaggatgaaggaggtgctc
 ggtgacaagggtgagaagggtgtgtctctgaccgtgtggtggactcctctgctgtctagtcactggggagtaggtgactgctaactg
 gagaggatcatgaaggccaggctctgagggactccagcatggccggctacatgtctagcaagaagaccatggagatcaacccggaga
 atgccatcatggacgagctccgcaagcgtgccgatgcggacaagaatgacaagtcctgaaggacctggatgctgctcttcgagactg
 ccttctgacctccggttgcagcttgaggaccccaacaccttggcaccaggatccaccgcattgctcaagctcggcctgagcatcgacga
 ggacgagtcgccgagggtgacgccgacatgccgccgtggaggacgacgccggcgagagcaagatggaggaggtcgactaa

Figure 11
 Page 7 of 7

Figure 12 (Example V)

>12464 OsGF14-c

atgtctcgggaggagaatgtctacatggccaagctggccgagcaggctgaaaggatgaggagatggtgagtacatggagaagggtgca
 aagactgtagatgtgaagagctcactgttgaggagcgcaacctctgtctgttactacaagaatgtgattggtgcccgcctcctgg
 cgtattgtctcatccatgaacagaaggaggagggctgtggcaatgaggaacatgttactctgatcaaggagtaccgtggcaagattgaagc
 tgagctgagcaagatttgcgatggtatcctgaagttgcttactcacaccttgtccctcatctactgtcagaatctaagggttttacctcaa
 gatgaagggtgattaccacaggtaccttgcggaatttaagactggtgccgagagaaaggaaagctgctgagagcacaatggtggcttaca
 ggctgctcaggatattgctctggcgatcttgcctccacccatcccataaggcttggactggcacttaacttctgtgttctactacgagattct
 aaactctccagacaaggcttgcacaccttgctaagcaggcggttgacgaagccatctccgagttggataccctcggggaggagcttacaag
 gacagcactttgatcatgacgctcctgagggacaacttgacctctggacctctgacctcacggaggacgggtggtgatgaggtgaaagaag
 cctccaaggcgacgcctgcgagggccagtaa

>20251 OsDAD1

atgccgagggccacgagcgaatgcgaagcttctaattcagtccttgggcaaggcttatgctgccacaccaacaatctcaagatcattgacct
 ctatgtggtctttgacgtgcacctgcccattatcaggtgtttacatgggaatagtggggtcatttcccttcaactcttcccttctggcgtcctctc
 atgcataggaactgcagtcctcgtgtttgccttcgtattcaagttaacaaagataacaaggaattcaaggatcttccctcctgaaaggcggttg
 cagatttctgtgtgcaatctggtgtcctcacttggatcatgaacttccctggttga

>19842 OsRCA1

atggctgctgccttctcctccaccgttgagctccggcgtccactccgaccaacttctggggaagaagctgaagaagcaggtgacatcgg
 cggtaactaccatggcaagagctccaacatcaacaggttcaaggatggccaaggagctggacgagggcaagcagaccgaccagga
 caggtggaagggtctcgcctacgacatctccgatgaccagcaggacatcaccagggggaagggttctcgtcactccctttccaggctccc
 acgggtgatggcaccacgagggcgtccctcagctcctacgagtacctcagccagggtctcagaacgtacgacttcgacaacaccatggga
 ggcttctacatcgccttctgcttcatggacaagctcgtcgtccacatctccaagaacttcatgacctcccccaacatcaaggtcccactcatcc
 tgggtatctggggagggaagggtcagggaaaatccttccaggtgtgagctcgtcttcgccaagatggggatcaaccccatcatgatgagcgc
 cggagagctggagagcggcaacccggagagccggcgaagctgatcaggcagcggtaccgtgaggcggcagacatcatcaagaagg
 ggaagatgtgctgcctcttcatcaacgatctggacgcgggtgcaggtcgcattgggaggcaccaccagtagacgggtgaacaaccagatg
 gtgaacgccaccctgatgaacatcgccgacaaccaacacgtgcagctccccgggatgtacaacaaggaggacaacccccgtgtccc
 catcatcgtcaccggcaacgacttctccacgtgtacgcgcgcctcctcgtacgggcgtatggagaagttctactgggtcccaccgc
 gacgaccgtgtcggcgtctgcaagggtatcttcgcaccgacaacgtccccgacgaggacatcgtcaagatcgtcagagcttccaggc
 caatccatcgatttctcggcgtcttctgcccgtgtttacgacgagaggtgcgaagtggtgtcggacacgggtgtggagaacattgg
 caagaggctggtgaactcagggaggggccaccggagttcagcagcccaagatgacgatcgaaggtcatggagtacggatacatg
 cttgtgaaggagcaggagaacgtcaagcgtgtgcagctggctgagcagtagctgagcagggctgcttgggtgacgtaactccgacgcc
 atgaagactggttcttctacgggcaaggagcacagcaagcaggttaacctgctgtgccggaagggtgcaccgacctgttgccaagaac
 ttcgaccaacggcgaggagcgacgacggcagctgcctttacacctttaa

>19902 OsEXPB2

atggctggggcctctgccaaggtcgtcgcgatgctgctcctcgtgctcgcacgtacggcttcgccgccggcgtcgtctacccaacgact
 ggctcccggccaaggccacctgtgtacggccagcccaacggcgccggaccgacgacaacggcggtgctgctgggttcaagaacacc
 aaccagtaccggttcatgtccatgacctcctgcggcaacgagcctctgttccaggacggcaagggtgtggtgcctgctaccagatacgg
 gcaccaacaacccgtcgtgctccgggcagcccaggacgggtgatcatcacggacatgaactactaccctgtggccaggtaccattcgac
 ctgagcggcacggcgttcggcgccatggcgaggccggggtgaacgaccagctccgccacgccggcatcatgacatccagttcaggc
 gcgtcccgtgtaccaccggcctctacgtgaacttccacgtcaggccgggtccaacccgggtgtacctcgcctgctggtggagttcgc
 caacaaggacggcacgggtggtgcagctcgcagtcagtcgctccccagcggcaagccgacgcgggtctggacgcccattgcgccg
 ctctggggatccatctggcgctcgcagccaaccaccgctccaggggcccttmtccctccgcatggtcagcgagtcggccagaccg
 tcatcgccaccagggtcatccggccaactggaggggccaaccaactacggctccaaagtccagttccgtga

>22832 OsBAA02730

atggcgtctgtacttctcctcaagtcatttcttcccaagaagctgaatgggggtgccaccgccagggtgccgcccccaagccggtgac
 tgtctccatggtgtccgcgtggtgctacgacgatgagctgtcaagaccgcgaaccattgcatcaccaggaagggtatccttcca

Figure 12

tggatgagtcgaacgcgacctgcggtaagaggcttgcgtcaaltggccttgagaacaccgaggccaaccgccaggcttaccggacctcc
 ttgtcaccgcaccgggcttgggacagtacatctccgggtgctatcctctcaggagacctgtaccagtcaactgtatagggaagattg
 tcgacatcctcactgagcagaaaatcggtccaggtatcaagggtcgacaagggtcttgtgcccttgtggtccaacaacgagtcattggtgc
 caagggtcgcgacggccttgcctcgcgcgaggcggcatactaccagcaggggcgctcgttcgccaaagtggcgactgtgtcagcatccc
 caacggcccatctgaactcgcgtgaaggagggtgcctggggccttggccgtacgccatttctcaggacaacgggctggtgccgat
 tgtcgagcctgagatcctcctgacgggtgagcatggcatcgacaggaccttcgaggtggcgagaaagtgtgggaggagaccttctctac
 atggccgagaaatgtgatgttcgagggcacctcctcaagccaagcatgggtgacacccgggtgccgagtgcaaggacaggggccacccc
 tgagcaagtatctgactacacctcaagctcctccacagaaggatccccctgccgtccccgcgatcatgttctgtcgggtgggcagtcgg
 aggtggaggcgacgcagaacctgaacgcgatgaaccagggggcccaacccgtggcacgtgtcgttctgtacgcgaggggcgctgcaga
 acacgtgcctcaagacgtggggcgggcagccgggagaacgtgaaggcggcgagggaccgctgctcctccgcgcaaggccaactcg
 ctggcgagctcggcaagtacaccagcgacggcgaggccgccgaggccaaggaggggcatgttcgtaagaactacgtctactaa
 >22840 OsAAB46718

atgatcacctcggcgtgggtggcgccggccgggccaaggggccttccgtccatctcccggcggggatcciccttcgccatcgtctgcagc
 ggtgggaagaagatcaagaccgacaagccctacgggattggagggtggcatgtcagtcgatattgatgcattcggcaggaaaagcacggg
 gaagggtgtgtaccaattcgtgacaagtacggcgcaaacgtcgacgggtacagcccaatctactcggcgagggaatggtctccaccgg
 tgacacctacgttgggtgaaccaccgggcttctgatctgggcccgtgacctcgcggggtcctcggcgggcgccctcctcgtctacaac
 accagcgccctcggcgctaa
 >22844 OsBAB61062

atggcgccaacgccggggtgcggcgggcggtgtccctggaccaggccgtggcgggcgctggcgggcggttctgtcgcggaagcagctgc
 gggtgcccgccggcgcgcggggggatgcgggtgcgggtgcggcgcgggggcgggcgggaggcggtggtgggtgcgcgt
 cgtcgtcgtcgggtggcagcgccggcgggcggaaggcgaggagatcgtcctcagcccatcaggagatctccggggcggttcagctgcc
 aggggtcaagtgcgtctccaacaggtacctcctcctcctccgccctcctcaggggcacaacagtggtggacaactgtctgaacagtgggat
 gttcactacatgcttgaggccctgaaagccctcgggctcctgtggaagcagataaagtgcacaaagagctgtagtcgttggctgtggtgg
 caagtttctgttgagaaggatgcgaaagaggaagtgaactcttctggggaacgctggaactgcaatgcgaccattgacagcagcogtg
 actgctgctgggtggaatgcaacttatgtccttgatggagtgcacgaatgaggagagaccgattggtgacttgggtgcgggtgaaaca
 acttgggtgcggatgtcactgttcttggcactgaatgccacctgttcgtgtcaagggaattggaggacttctggtggcaaggttaagctc
 tctggttccatcagcagtcagtacttgatgccttgcgtgatggctgtccttggcccttggggatgtggagatcgaaatcattgacaaactaat
 ctccattccttactgtgaaatgacattgagattgatggagcgttttgggtgaaggcgagcattctgatagttgggacagattctatftaaggg
 agggcagaagtacaaatccttgaaatgcctatgtgaagggtgatgcctcaagcgcgagctatttcttggctggtgctgcaatcactggag
 gcactgtgacagttcaagggtgtgtgtacgaccagtttgagggtgatgtcaaatgtctgaggtacttgatgatgggagcaaaaggttacat
 ggactgacaccagtgaaccgtactgttccaccagtgagccttatgggaagaacacctgaaagctgttgatgtcaacatgaacaaaat
 gcctgatgttgccatgaccttgccgttgggtgactctcgtgatgggtcactgctatcagagatgtggcttcttgagagtaaaaggaaacc
 gaaaggatggttgcaattcggaccgagtaacaaagctgggagcatcggtgaagaaggtcctgactactgcatcatccccaccggag
 aagctgaacatcacggcaatcgacacctacgatgatcacaggatggccatggccttctccctcgtgcctgcgcccagctgcccgtgacg
 atcagggaccttgggtgcacccgcaagaccttcccaactacttcgacgttctaagcacttctcaggaactga
 >22858 OsPN22858

gcttcttccggactgttgggtgctaaaatcactcaggaaactggtgatttctttagcagatgcagagggtgacccagacaaaccaactgatg
 gttttctctattgatagggtataggcgcatgcatgaaggaaagttgtattgtgtatgatgaaagcggtgataatgaaggggatcttg
 tcatggcagctacgttggcagaccagaatctattgcattcatgatcaggaatggttctgggatcatctcagtgggcatgaaggaaggagac
 ttaacaagattgatattcctatgatgtcctcaattgcagaaattgaggatattcagctgctgctccacagtaacagtggatgccagagtggg
 catatcaaccggcgctcggctgcagatagggcaaaaacgattttactctagcctcccctgattctaagccaactgacctcagaagaccag
 gccatatattcccttaaaaataccgaaacgggtggtgtgctaaaaagagctggacatactgaggcatccgctgatcttgcgcgttgggtgctt
 gcgcctgtgtctgtcgtcaaccgtcatcaaccagtggtggttcaatggcaggaatgccagtgtgaacagatggcttggagcatg
 atatccaattgttcaatcgtgatcicacccggtatagaagggaaggaggagaagctggtggaactgattgctgtatctcgttggcgacgaa
 atggggcctttccgagcttactgctaccaatccaagcttgatggaaccgagcacattgctgttgcacaggggcgacatcgggcgacggcgag
 gacgtcttgggtgagggtccattccgagtgcctgaccggcgacatcctcgggtccgcccgtgcgactgcggcaaccagctggacctggcg

Figure 12
 Page 2 of 5

atgcagctcatcgacaaggccggccgcccgtctctaccccgccggccacgagggccgcccgcctcggccagaagctcc
 gcgcctacaacctccaggacgacggccacgacaccgtccaggccaacgtcagctcggcctcggcgtcactcccgcgagtacggcat
 cggcgcccagattctcggggacatgggggtgcgcacgatcgggtgatgacgaacaacccggcaagttcgtcgggctcaagggtac
 gggctcggcgtcgtcggcagggttcgggtgatctccccatcaccaaggagaaccagaggtacctcagacgaagcgaccaagatgg
 gccatgtctacggctccgacctccccggcaacgtccggaggaattctcaacccggacgacatcgccggagaccaagacgaagacga
 caccacaactga
 >22866 OsPN22866
 ccacgcgtccggagaagcggccgctttttttttttttccctccgactccgactccgatcgtatccaccccgaatccctcctcaccgc
 cggcgagcttccggctcggcgagatggccacacgtattggatcgtctcttcccggtcagactcctggctccaccgccaattctctg
 ggcgcgctccaggactccatctcgcgccactcttcgacacgccgtctaccggttaacgtccccgatctccgcgtcggcagctcgac
 tccctcctcggcctcagcgacatctcgtcaagtccaacgtctcatcgaggggtctcgcacaagatccggaggcagatcgaggagctag
 agcgcgccgggggtgtcgagagtggggctctcaccgttgacggcggtccccgtcgacacctaccctaccaggtttgtgtgggatgagggc
 aaatacccaacgatgtcaccgctcaaggagattgccggcagcatccaatcacagggtctcaagattgaagatgacatgaaggttcgagga
 gcggaatacaataatgtaaggagccagcttaatgcgataacagaagcaaacgggaagcttagcagttcgtgatcttccaatctgtgtaa
 accagaggatattggtcacatcagaacatctagtacactccttgagttgttctcagttactctcaaaaggattggtgtcaagctatgagtccc
 ttgacacattgtgtgtaccgagatcgtctaaaaaactttatgaggacaatgagttgctctctacacggtaacattgtttgtaaggtgttgaca
 actttaaggctcgtcacgtgaaaaagggttccagggttcgcgattttgagtatagttctgaagcacaggaaagtaggaaggaagagctggaa
 aagctaagcaagaccaggaagcaatgagggcattcacttctgcaatggtgctatgccagctacagttaggtattcagttcctggatgcacttc
 tgtctggtgcgtgtcttttagagagcattcttagataatggtcttccccatcattctgtctgttctagcaccttctcaaaagggtgaaaaga
 aagtaaggagcatctgaggaactctgtgggcaatgtccatagtatttattgaaatctgaagacgatgttggtgtagctggtctgggaggca
 aggcagtgttttggagtaa
 >22874 OsPN22874
 ggaacaaacctggcttcagagttggagagatcaggctctccaacagggatatttatttcggcacattattggggaacacaccagagggttc
 agggaggtatgtctggtcagatggttgacttacgatggtgagtgaggagaggatgaggcatgggcaaggaaagacaatgtggccat
 ctggagccacctacgaggggtgagtactctggtggctacatttattggtgaaggcacatataccgggtctgacaacatcgtctacaag
 >23022 OsPN23022
 aagtatgctgagcgtgggcttctgtcttctgtgttgaagacagggaagtacgtgagaagtcgaaggagtctgctggtggaccatggcaattc
 gttggtctgctgcccctgtttgatccccgaggcacgacagtgctgaaactatccggaaggctctccatcttggtttaatgttaagatgatca
 ctggtgaccagcttgctattgtaaggagactggttaggaggttggaaatgggtacaaacatgtatccttcacccgattgctcggacaaaaca
 aggacgcttacttgaggcacttcccgtggatgagctcattgagaaggctgatggtttgccggagtcttccctgagcacaatatgagatcg
 tgaagagggtgcaagagaagaagcacatcgtcggatgaccggagatgggtgcaatgatgccctgctcttaagaaggccgacattggtatt
 gctgttctgatgcaatagatgctgcaagaagtgtccgacattgtgctcactgagccaggtcttagtgctattatcagcgtgtcctcactag
 cagatgcatttccagaggatgaagaactataccattatgcagtttccatcaccatccgtatagtgttggattttgttattgtctgatctgga
 aatacgaacttccaccttcatggtccttattgcatcctgaatgacgggtactatcatgacgatctcgaaggacagagtttaagccttctccct
 tgcagacagctggaagctgaaggaaatcttgcctaccggtattgtgcttgaagctaccttgctctgatgaccgtattttcttctgggccaatg
 cacaagacagactttttcacggacaaattcgggtgtcagatcaatcaggaacagtgaacatgagatgatgtctgactgtacctcaagtcagt
 attgtgagtcaggccctcatcttgcacccgttctcgcagctggcttcatcgaacgccctggttcttcttggctactgcattcatgcttga
 caactggttgcgacaltctcgtgtctatgccaactggggccttgcaggatcaagggaattggctggggctgggctggcgttatctgcttt
 acagcattgtgttacttccctcttgacatctcaagttcttcatccgcttgtgctgagtggaagggtcgggacaacctccttgagaacaaga
 ttgcaatccaccaagaaggattacggtagggagggagggagggcacaatgggccaccgcacaaagaacacttcacggccttcagcca
 ccggaggtcgcctccaacacgcttccaacgacaagagcagctaccgcgagctcttgagatcgctgagcaagccaagagacgagctga
 gatcgagggctgagagagctgaacacttcaaggggccacgtcgaatcggtggtgaagctgaaggggctggacatcgacacgatccag
 cagaactacacagtgta
 >23053 OsPN23053
 tcatccttcttggggttatctggtgtaccaataattggtggagcactggctgactactatggcggaagcgagttatggcttatggcgtggc
 tctatggtccttggctacattccttcccccttgggcagctgctcgctctctggtgttctctcaactagattctgctgggtatggcagaagg

Figure 12
 Page 3 of 5

gggtggcactcccaagatgaacaacatgggtgtgaggtggttctcgcacagagcgatcaagtgtgtcgggattgcaatggctggtttca
 gcttggcaataaccattggtttacttcttccctatcatcatgtcacagagctggaattttggaccatttgtgatatttggattttgatgg
 gtgttggtgtggatatctgctatatcagggactccaggtgaaaatgcacaaatcagcacatgaactagattatataaccaggggtcagaaa
 ttgtaaaaactcaatctggaggtgaaagattacgaaaggctccctccttcagcaagctactctctaaatggccaacatgggctttaattctgc
 aaatgctatgcatagttggggttatttgcattcttcatggatgccagtgtatttcaaaacaatatatcatgttaatctgagagaagctgcatggt
 tagtgactcccctgggtcatgatggcagtttaggctatgttctgtgtgtgtatcagacaggcttatccaaaatggtacaagcatcacttgac
 tcggaagataatgcagacaattggcttgggtcctgggtgtgcttcttggactaaatgcagcaaagagtcagtcacgttctgtctgg
 ctactatcgctgtaggttgaagtccttgggtcactcaggtttctagtaaactctcaggagattgtccacaatatgtggagtgtacatggaa
 tgtcaaatacagctggaacatttctgcccatttaggaactgttgagcaggttcttctgcatcgatgggttcttccgtggatttttaataata
 catctcttatacttcagtagcacactgttctgggataatttgcactggagagcgtgttgatttgcactggcactggtag

>23059 OsPN23059

atggcagcatcgctccaagccgcgccaccctgatgcagccggccaagctcggcgccgggctcctccgccgcgctgccatcgccgc
 cgtcttcgcacgtcgccaggcggttcggcgctgacaccggtgcccggcgagatcacgtgctccctgcagtcggacatccgggaggtc
 gccacaagtgcgccgatgccccaagctcggcgcttcgcccctgccacctcagctctgctcgtcggcgccagcgcgaggggcgt
 gccgaggaggcttaccttcgacgagattcagagtaagacgtacatggaggtgaaggaaccggcacggcgaaaccagtcccgcaggggtg
 gaggggcgcgctgactccttcgcttcaaggccggcaagtacaacatgaagaagtctgcctggagccgacgtcctcaccgttaaggcg
 gaggggcggtggccaagaacgcgccaccgagttccagaagaccaagctcatgacctgcctgacctacacctcgcagagatcgaggggc
 cgctcgaggtcagctccgacggaaccatcaagttcgaggagaaggacggaatcgactacgccgccgtcaccgtgcagctgccgggagg
 cgagcgcggtgccgttctcctaccatcaagaatctgtgcgccaccggcaagccggagagcttcggcgggccctcctcgtgccagctac
 cgtggttcgtccttctcgacccaaaggcgcggtggctctaccggctacgacaacgtgtggtcgctccccgccggaggcagaggaga
 cgaggaggagctcgtaaggagaacgtcaagaacgctcgtcgtcgcagggcaacatcacgttgagcgtcaccagagcaagccagag
 accggcgaggtgattggcgcttgcagagcgtgcagccgtcggacacagacctcggcgccaagggtgcccaaggatgtcaagatccaag
 ggggtgtgtacgcgcagctcagtag

>23061 OsPN23061

atggccatggccacgcaagcctccgccccaagtgccacctcctcgccgcctgggcaccggcggaagccgcgctcatccacctctccat
 gccacctcgaggcgaccacacctccctcagagcgggcgggaggatcagcccggcgggcgacggaggagaagaagccagcc
 cccgcggggttcgtgccgccagctggacccaacacgccgtccccgatcttcggcgggagcacgggggactcctccggaaggcg
 caggtggaggagtctacgtcatcacatggacgtcgccaaggagcaggtgttcgagatgccacggcgggcgcccatcatgcgcga
 gggcccaacctgctgaagctggccaggaaggagcagtgctggccctgggcaccagggtccgctccaagtacaagatcaactaccagt
 tctaccgcgtcttcccaatggcgaggtgcagtacctccacccaaggacggcgcttaccgggagaagggtcaacgccggcaggcaggg
 cgtcgccagaacttccgcagcatcggaagaacgtcagccccatcgaggtcaagttcaccggcaagaacgtc
 ttcacatctag

>23426 OsRBCL

atgtcaccacaaacagaaactaaagcaagtgttgatttaaagctggtgttaaggattataaattgacttactacacccggagtacgaacc
 aaggacactgatatcttggcagattccgagtaactcctcagccgggggttccgccgaagaagcaggggtgtagtgcggaatctt
 ctactggtacatggacaactgttggactgatggactaccagcttgcgttacaaaggccgatgctatcacatcgagcccgttgttggga
 ggataatcaatatatcgcttatgtagcttatccattagacctattgaagaggggttctgttactaacatgttacttccattgtgggtaacgtattgg
 ttcaaagccctacgcgtctacgtctggaggatctgcgaattccccctacttattcaaaaacttccaaggtccgcctcatggtatccaagttg
 aaagggataagttgaacaaatacggctgctcttattgggatgtactattaaacaaaattgggattatctgaaaaaattatggtagagcatgtt
 atgagtgtctacgcgggtggacttgatttaccaaagatgatgaaacgtaaaactcacaaccttattgcgttggaggggaccgtttgtctttgtg
 ccgaagctatttataaatcacaggccgaaaccggtgaaattaaggggcattacttgatgcgactgcaggtacatgcgaagaatgattaaa
 agagctgtatttgcgagggaattaggggttctattgtaatgcagtacttaaccgggggttcaccgcaataactagtttggtcattatt
 gccgcgacaacggcctacttcttccattaccgagcaatgcagtattgatagacagaaaaatcatggtatgcatttccgtgtattagct
 aaagcattgcgtatgtctgggggagatcatatccacgctgttacagtagtaggtaagtagaagggaacgcgaatgactttaggtttgtt
 gatttattgcgcgatgatttattgaaaaagatcgtcgtcgcggtatcttttactcaggactgggtatccatgccaggtgttataccggtggtt
 caggggggtattcatgttggcatatgccagctctgaccgaaatcttggagatgattctgtattgcaatttgggtgggaactttaggacatcctt

Figure 12
 Page 4 of 5

ggggtaatgcacctgggtgcagcagctaatacgggtggcttagaagcctgtgtacaagctcgtaacgaaggcgcatcttgctcgtgaagg
 taatgaaattatccgatcagcttcaaatggagtcctgaactagccgcagcttgtgaaatattgaaagcgatcaaattcagttcgagccggg
 agataaactagatagctag
 >29982 OsPN29982

gcgccagggcgcgagtgctcctggcacggccgctcctccgccgctcggcgccggggcgctgcctgcgcccttcgccggaggccctcca
 agtacaagaataaaatccagaatgaggaggttagtagaggatgatattggcgggtggtggtgaagacgatgacgacgcgtggaagcac
 ttttaagcagttggaagaagatcacaagaatgatgattgtctgttaggatgatgacgatggaatttcagaagaagatatggcaagatttgag
 caggagttggcagaggcaatcggggatattgtctgatgctgatgaatctggagagggttcgtcattaggctctgaggcttatgggaatgatga
 aaaaacggatgaaatcaaacggccagagctgaaaaactggcaactaagaggctggctcgtcctcaaaaataggctgccgtaaaaactagt
 ataaagaatcttgcaggggaactaggcctggataggacttgggtcattgaattgctccgaatccacctccaaaactctattcatgtctgattct
 ttgcctgatgaagaccttctaaacctgaaatcaaggaaatagagccctcctcagtagttgataacgctgatgtcactgaaaccaagccacag
 acggaacttctgtcatgtatgtgcgcagaatggtcttcacagaaaagattaaaaaagggtgcaactggagacattagaaagagtctactcc
 cgaaccaaacgccctacaaatacaatgatcagcagcatagttcaagtacaagccttcacggaagaccattgttaagtgggttgaggatag
 aagggaacaggatgggggtacctgaccatcgagttgcatcaagagatccctatctgagaccatagctagtt
 >30846 OsPN30846

gccccacggcccccctccaccttctcaacgccgtcgcgctcggcaacgtcggggcggggcaagtcggccgtgctgaacagccctcatcgg
 ccacccgtgctgccgacgggggagaacggggcgacgagggcgccaattgtgtggacctgcagaggggaccgggcccctcagcagca
 agtccatcgtgctgcaaatcgacagcaagtcgcagcgggtgtccgcaagttcgtgcggcattcgtgcaggacagggtcagcaagggg
 gcgtccagcggatccagtagggggcgtgtagaggggatcaatcacaagctgcggacgagtagcagctccccactaaagctggttgattac
 ctgggatagaccagcgagctgtgacgatccaatgattaatgaatatgctgggcacaatgatgaatactgctagttgtgatacctgcaatgc
 aagcagctgatgtgctcatctcgagctcttaggctggccaaggatattgatgcagatggcaccagaactgtaggtgtgataagtaaagttg
 atcaagcagaaggagatgcaaaaactatagcttgtgttcagggcccttctgttgaataaggggcccaaaaaaccttcagatatcgagtggttg
 ctctgataggacaatcagttgcaattgcgtcagctcaagcagcaggttctgaaaactcactagaaacagcatggaatgctgaagctgaaacc
 ctcatatccatcttaactggagctccaaaaagtaaaactaggtagaatagctctggttgacaccattgctaagc
 >30974 OsPN30974

atggggggcccttttagctcccccaacagcaagaaccagccatgggagcatggtgaagcttcaaaggcagattcctccaagaagctgcgg
 atgtcggcgccgctctgtccggtggtgtacgaccacccgggctgatccgggggctcccgatgagatctcgtcagatcctcgaagg
 atgccccggatgggttatctgaatgcaaaagatggttcgaggagctggaaggctgcgatcactggcgtggagctgtaccgggtgaggaag
 gagctgggtgtgagtgaggaatggctgtacatgcttaccagtcagatgatgggaagctagtgtggaatgcatcgcggtttgtggcca
 atggcagaggctgccactgatgccggggatcagccatggaggagaatgcaagaggggcatccctgggttgggttaggggatttgtt
 agcggccggcatcagggtctctgatgttattagagctggcttggccaaagggaattcattggataggcttccgttctgcggttgcgaattggg
 acagtcaacgggtgcatctatgttttaggtggattctctagaggctcggcgatgaaatgtgtatggaggtatgatcctttgtcaatgcgtggca
 ggagggttagctcgtatgagcaccggcgcgcttctgcaagggtagcctgctgaacaagaagctgtatgttgggtggtgtcagcaaaaggc
 aagaacgggttagctccgctccaatccggcaggtgtttagcccaaggacaggaatttgggtgggagggggccctgacactctccgtctc
 aaaggggcaagctctaccagctgcctcttgggtgagctggtgaagccattgcgacaggaatgacctcttgggaggcaagctttatgttctt
 caaagctgtattccggggccattcttgggtgatgttgggtggggagatcttgcgagacaaaattcatggcggaatgcctgtaggaatgg
 gcgagggtggtcagcagggcaggtggttacgaagtaagtgtgtcatagatggggacttgtatgcttggagccatcaactcttttgata
 gaggtgaagatcaagattatgatcctcaggaggtgcttggaaaggttgcattggccaggtaccagtggtgacttgcagagtcggaatgt
 ccatacttacttgcaggatttcttggcaagctcaattgatcatcaaaagatgtagatagcaagatcaatataatgcaaacgatgttctgaagcct
 gtggagttgtcagcccctggaaatggtccaacatgccaaaatcaacaactttttcagaacaggaaactaatttggaaagtcatcgtgtcga
 aaaaacttgcagccgctgaattagtcagttgtcaggtgtcaacatat

Figure 12
 Page 5 of 5

gatcgccacccagacttccaagatgatccttacatctacgcctttgacagcctgaagtccattggaattgagttgtggcaggtcaaatcagg
gactctgttgacaacattctgattactgacgatgctgctttggccaagacattcgagaggagacatgggccaagcacaaggatgctgaga
aggctgcttttgacgaggcagagaaaaaaggaagaggaggaaagctgccaaggccggtgaggacgatgatattggatgacgagga
tgctgaggatgaggacaaggccgatgagaaggctgactctgatgcagaggatggcaaggattctgatgatgagaagcatgatgagctcta
a

>20551 Os003118-3674

atgggggggaaggagctgagcgaggagcaggtggcgtcgcgatcggggaggcgttctccctcttcgacaccgacggcgacggccggtac
gcgccgtcggagctcggcgtcctgatgcgtccctcgccgggaacccacccaggcgagctccgcgacatcgccgcgagggagaag
ctcaccgcgcccttcgacttcccgcttcctcgacctcatgcgcgccacctccgccccgagcccttcgaccgcccgtccgcgacgcct
tcgcgtcctcgacaaggacgcctccggcaccgtctccgtcgccgatctccgccacgtcctcactccatcgccgagaagctcgagcccc
acgagttcgacgagtggatccgcgaggtcgacgtcgccccgacggcaccatccgctacgacgacttcatccgcccgtcgtcgccaaat
aa

>20554 OsRAB16B

atggagaactaccaggggcagcacggctacggcgccgaccgcgtcgacgtgtacggcaacccggctggcgccggccagttacggcgg
cggcgccaccgcgcccggcgaggccacggagcgcgatggggatgggaggtcatgccggcgccggcgccggcgccagttccagccg
gcgagggaggaccgcaagaccggcgccatctccaccgctccggcagctcaagctccagctcgtcgtctgaggacgacgggatggga
gggaggaggaagaaggggatcaaggagaagatcaaggagaagctgcccggcggaacaagggaacaaccagcagcagcagcag
atgatgggggaacactggcgccgcgtacgggcagcagggccacgccgggatgaccggcgccggcaccggcggtgcacgggtcgaggatg
cggcaacgccggcgagaagaagggaattcatggacaagatcaaggagaagctgcccggccagcactaa

>22883 OsLIP5

atggccggcatcaccacaagatcgaggagaagctccacatggcgaggcgagcacaagaaggaagacgagcacaagaaggagggg
ggagcaccacaagaaggacggggagcacaaggaaggcgtggtggagaagatcaaggacaagatcaccggcgaccacggcgacggc
ggcgagcacaaggagaagaaggacaagaagaagaaggagaagaagcacggcgaggagggtaccaccacgacggccacagc
agcagcagcagcgacagcgactga

>23226 OsPN23226

ataaaaagacttcaccttctgctcacagtgaaggaatctgctatggatgttctacaaacctggatgctagaaggcgatattctttgctaa
ttctcttttcatggacatgccaagtgtccaaaagtcggcacatgttgccttctctgttcttactccttactacaaagaagatgctctttctctt
ccaagcattagaagaccagaatgaagatggggttctattcttttacttgcaaaaaatctatccagatgaatggaacatttcttcaaagggt
ggattgcaatactgaagaggaaactcgtgagacagagcagttggaagatgagcttcgctttggcgcacataggggccaactttgacg
agaactgtcagagggatgatgtactacagacaagctttggtgcttcaggctttttgatatggctagagatgaagatttagggaaggcttca
gagcagctgacctttaaattgatgaatcaccattactgactcaatgcaagctatagctgacatgaagtttaccatagttgtatcatgccaa
tacggtatccagaacgttctggtgatcaccgtgcacaagatattcttagactgatgacaacttatccatcacttcgggttgcttatattgatgaa
gttgaagagccaagcaaagacaggaacaagaagatagaaaaggttactactagcgttggtgaaggcagctgtaaccaagcctgacgat
cctggtcgaaaacttgatcaggacatatacagaataaagctaccaggtaatgcaatgctagggtgaaggaaagccagaaaatcagaacctg
cgataatatttactcgaggtgaaggccttcaaactatagacatgaatcaggaacattacatggaggagactttgaaaatgagaaatctgctgc
aagagtttctgaagaaacatgatggtgtgaggtatccatcaatacttgggtgtgagagagcacaatttactggcagtggttcttcttgcgtggt
tcatgtcaaaicaagagacaagttttgctactattggacaacgggtacttgcgaatcctttgagagttcgatttcattatggacatcctgatatttt
gatcgacttttccacctcacgaggggtggtgtaagcaaagcatccaagattatcaatcttagtgaggacatatttctggttcaactcaact
gcgtgaaggaaatgttacgcatcatgaatacatgcaagttggcaaggggagagatgtgggtctcaatcaaatctcactatttgaggcaaaaa
tagctaattggaatggagagcagacactgagccgtgacgtctaccgacttgacatcgtttgatttctcagaatgttgctctactacacc
actattggttttacttcagcactatgatgacagtggtgacagtgatgtattctctatggacgtctctatctgtcctcagtggtcttgatgaagc
cttggctactggaaaaaggtttatacacaatgaacctctccaggtgcttctgttcacagctctttgtgcagctgggttttgatggcgttgctta
tgatgatggaaattggtctggagagaggatttagaaccgattgagcgactttgtactgatgcagctgacgttagcatctgttttctt
tacattctctctcgggacaaaactcactactatggaacaacgtgctccatggaggagccgagtatagagctactggcggtggaattgtggt
gttccatgccaatttgcggagaactatcgactatactacgcagccattttgcaagggtattgagttgctgattttgctaattgtgtatgaaatc
tttgggcaatcatatcgaggagctatcgctacatcttattacatttctgatgtggttatggtgtaacctggctcttgcaccattcctatttaac

Figure 13
Page 2 of 5

cttctgggttcgagtgagcagaagattgtggaatgattggactgattggaataagtggatcagcaaccgtggtggtattggtgtaccaccggaga
 aaagttgggagtcagtggtgggagaagagcaggagcctataaaatattctgggaagcgtggaattgttctcgaaatagtgcttgattgcgct
 tcttatctaccaatatggtctgtttatcactgaacataaccaaacacacaaaagagtgctcctggtctattgctgtcatgggtgtcatctttaa
 ttctgcttgatgaagaccgtgtcagttgggagcggaattcagcgcggactccaacttggtccggttgattaaggggtgatattata
 acatttatctccatcattataatcttgatgcaatccctcatatgacagtccaggacatcttgttgcattctgccttcacccaactggctggg
 tctgctgttggtgcgaagctatcaagccagtaattgtgcgcatcggtgtggtggatcgaatcaagcgcttgaggggatgatgatcat
 catggggctcctgctctcaccgccatcgcttccctgcttggtcccggttgatccgagttccagaccagaatgccttcaaccaggccttca
 gcagaggtctgcagatctgcgaatcctggcgccacaagaagaccgagctacacggaacaaggagtag
 >23485 OsPN23485

cacgcgtccgggtccgggtcgggtgcgggagcgggtcaggccatgatccgccgctgcagggggaggtgtgcgcgcgctggagga
 ggccgacgggagcggcgcccggttcgtggaggacgtgtgtcgcgccccggcgcggggggcatcagccgggtgctccaggac
 ggccgctgttcgagaaggccgggtcaacgtctccgctgtacggcgtcatgccccgacgctaccgcccgaagggggag
 ccggcaagaacggcgcgggcgagatggcccaaggctgggcccgtgcccttcttgcgctggcattagctcggttcttcccccaaga
 accattgtcccaacattgcatttaactaccgctatttgagacagatgcaccgaaagatgctcctggtgcaccaaggcaatggtggttgg
 aggtggtactgattgactccttcatatcattgaaggagatgtaagcatttcttctgttcaaaaacaagcgtgtgataagttgaccaagt
 tttaccggagattcaaaaaatggtgtgatgattattctatattaagcaccgtaatgagcgtcgggggcttggtggaatattttgatgatctaa
 tgaactatgatcaagaatgcttctcaacttgcacagaatgtgcggattctgttctcgcatacataccaatcattgaacgccggaaggaca
 ctccatttactgaggaacacaaggcatggcaacaactgaggagaggtcgctatgtggagtcaacctgtatgatcgcggtaccacattg
 gcctcaagactggggggaggagtgaagatattctggttcttccactgactgcacgatggcagttattatcacacctgaagagggaactg
 aagaacggaaacttctgacgcgtgcataaaccacaaggatggtcgtatctga
 >23878 OsPN23878

garargggtggaggtkacgcggaggtggattcggcggtgctgctactggtgctcgggccacggcgggcgacgacaccacggaggcag
 tctccggcgggagctgggagccctcgcagcgcgacgcgcgggtgggtgacacgagctcctacgggtgctagaagaagcacagggtt
 ttgccagcttccaaatagtgattgggccttatgactgttaattacaacctccggtgacgattccgactactaaagcttccgaaggaaagtactg
 aggttgaaaacagagagcctgaagctgcaaatcctgaaattcttgacggggtggatgatcttatgcagctaaagtatctaagtgaacatcc
 gtctgtacaatttgacgtacagatacttcaagacttgatataactaaagcaggtccagtttgggtggcggtcaatcctttaaagaagttcctc
 tatatggaatgagtagatgatcatatagaataaaacaaggatagccacacgtttatgcaatagcagattcagccctccgtgaaatgaa
 aagagatgaagtaaccagctcatcatataagtggtgagagtgagcaggaagacagaaactgcaagattgctatgcagtattggctt
 ctcttggaggtgggggtggcatagaatatgagatcctacaacacccaataactcgaggcttttggcaacgcaagacactaagaatga
 caactcaagtcgcttggaaagctcattgaaattcatttagtacaactggaaggatttgggtgccatgattcaaacatttacttgagaagtca
 aggttgtacaatgtgcagttggcgagcgtcctaccatattttatcaactatgtccgggtgctcctgcgtccctcagagataagttgaatatg
 aagaagcggatgaatacaaatattgaagcagagttgctgctattcaattgccggtgtggatgatgctcaaatgtccgtagtgaacggaag
 ctatgaacatcgctcatatcagcaagaggaccaagataatgtttcacgatggtttctgcaattctatggctaggagatgtccttccaggtca
 ttgacaacgaaaacatgttgagattgttagatgaagcggcagaacggctgcgaagacttcttggtgcagattgaagatctcaatttagc
 ttgtcgaagaggcacatgaaagtaataatgaaatattgtacagaaactcacccttccacaggcaatgcacaaagagatgcgttggccaa
 gtcactctatgccagtttgttgagttggttgaacaaattaacaagctcttccagttggcaagcgtcgaactgggagatcaatcagcattctt
 gatatctatggcttgaatcattcgataggaaacagtttgaacagttctgtattaactatgctaatgaaaggctccaacaacattcaatcgcattt
 gtttaagcttgaacaagagggaatagctgaagatggcattgactgggcaagggttgagttgaggacaatcagaactgttgaatctctttag
 aaaaagccactgggggtgtgtctctgtagatgaagaatctaccttcccaatgccacagatcttacttttgcgaacaagcttaagcaacatc
 gaacaataattctgcttccgaggagaaagaggcaaggcttttcagtcctgactatgcaggagaggtggcttatgacacgtcaggttttct
 agagaagaatagagatttattgcacatggactcaattcagttccttgcgaatgcaatcctcactaccacaatgttgcattcaaaatgcttcc
 tcaatctgataatccgttacctgttccatatagaatagtgtgctgactcacagaagtttaagttgcatgaagtttaagggaacattgttcca
 acttatgcaagactcgaagtagaacaccacacatttatcgttatttaagccaacaatttgaactccctgcaattatgaacaaggacttg
 tgcctcaacaactcaaatgttggtgggttcttgaggttgcgaattcaagatctggatataccaagaatgactcatcagaatgtgctgcac
 ggtatggcttcttctttaggagtgctgcacatcaagaccacttaggttccggttgaattctcaccagtttaacatttgcctgagatgatc
 aagttgggtacacaaaattgttcttgaactggccagattggcaaaactgaagacactcggaaccgtactctgcatggtattttaagggttcaa

Figure 13
 Page 3 of 5

agctgttcagagggcatcaagcccgtcgccatgcaaggggaacgaalaaggaggttttgctcttcaatcatttattcagagtgagaatgca
 agaaaaatgtattcgtctctggcaaggaaacacagggcagccattattttacagagaaatctaaaatgctggcttgaagaagatatttcgtta
 acatacgggaaggcttcagtagtaatacaaatctggcaltcgtggttgccttgcgaagatgtgctggcaatgttgatctattgaatgtgctaagag
 aattcgaatcaaagaaggagggcagaggggtgaccaaatttgattaaggcatcattcctggctgagctacaaaggcgaatattaaaggccga
 ggcgacagtcggagaaaaggatgaagagaacgagatgcttcagcagcgacttcagcaatatgaaaaccgatggtcagaatacagagcaga
 aaatgaaagcgatggaggaatgtggcagaagcaaatgcggtcattacaatcaagcctatccgtagcaaaagaagagccttgcttgcgatga
 gacgcctaggatgtccgactcatcggtcgaccagagctgggagagcaatgggggtcacattggcagcgcgctgcagctggttccacgcac
 cgttggacgggagatgaatgccagcatcagcgtcatcagccgcttggccgaagagttcagcagcgagccaggtcttcgctgatgatgc
 caagtcttgggtgaggtcaagtcggagcagggccgatgccagcctgaacctgacatggagctccgcaggtgaagcagaacttcgactc
 atggaagaaggacttcggctcgccataagggagacgaaggtgatcctcaacaagctggggagcggcaacgaatcgtcgcccaactcg
 gtgaagcgaatgtgggggagggctgaacacatccaagttttcatga
 >24059 OsAAK01712

atggacgcctccgcccggagggcgggggaactcgtgccgacggcggggggccgacggggccaagcggcggggtgtgccacttctacga
 cgcggaggtggggaactactactgcgggcaggggcacccgatgaagcccaccgcacccgatgaccacgcgctgctcggccactac
 ggcctcctcgaccagatgcaggtgctcaagccccacccggcgcgacccgcgacctctgccgttccacggcagcactacgtcgcctt
 cctccgctccgtcacggcgagacccagcaggaccagatccggcgctcaagcgttcaacgtcggcgaggactgccccgtcttcgacg
 gcctctacagcttctgccagactacgccgggggatccgtcggcgccgctcaagctcaaccacggccacgacatcgccatcaactgg
 gccggcgccctccaccacgccaagaagtgcgagggcctccggcttctgtacgtcaacgacatcgtcctgccatcctcagctcctcaaat
 accaccagcgtgttctctatgtggacatcgataccaccatggggatggtgtggaggaggcgttttacacgacggacaggggtgatgacggt
 ctggttccacaagtgtgggattatttccgggaacaggggatacctgtacattgggtattcagaggggaagtattactgcctgaatgtccc
 ctggatgatgggattgatgatgacagctaccagtcactcttcaagccgatcatcagaaagtcagagatgtatcgtcctggtgcagtcgtg
 ctacagtcggcgctgattcgttgcggcgataggttgggtgttcaatctctcagggaaggtcatcgagaatgtgtaaagttcatgaggt
 ctttcaatgttccgttgccttcttgggtggtggatgccataagaaatgttcacgctgttgggtgatgagacaggagttgcacttgggtca
 tgagctcaccgacaagatgcctccaaatgagtattttgagtattttgtccagaatacagctctttcgttcagcaagtaacatggagaacaga
 aataccaacaagcaattggaggaaataaaatgaacattctggacaatcttcaaaacttcaacatgctcctagtgtccaatggaagagcgaa
 ttctgaaacaaagtacctgagccagatgaagatcaagaggatccagatgaaaggcacgacctgactctgatattggtattggtatgcac
 aaacctatgggacactcagcaagaagccttattcgaacatcgaagttaagagagaaatcactgaatcagaggccaaagatcagcatggta
 agagattgacaactgaacataaaggaccagaaccgatggcagacgatcctggttctccaagcaagctcctgtaagtc
 gtcgtcttctatccatctgcaaatccatag
 >29037 OsPN29037

atggaggtggggttcttggggctgggcatcatggggaaggcaatggcggccaacctcctccgccacggcttccgcgtcaccgtctggaa
 ccggactcttccaagtgcagggagctcgtcgcgttggcgccgctgggggagacgccggcgccgtcgtcgccaagtgcagatac
 accatcgccatgctcctccgacccagcggcgctatctgttattcgacaaggacggcgtgctcgagcagattggggaagggaagggt
 tatgtggacatgtccactgttgatgccgaccttcttgaagataagcgagggtataaaacaaaagggtggggcctttgtgaagctccagtti
 caggaaagcaaaaaggcagctgaagatggccaattggtcatttctgtcaggggacaaggattgtatgatgatattgtccctgcattgatg
 tacttgggaaaaagtcgttcttttgggagagattgaaatggagcaaatgaaactggtgtcaacatgatcatgggaagtatgatgaatg
 ctttgcctgagggactctcttggctgataacagtggttgagccccagacacttcttgatgtcctggaccttggcgccatcgcaatccaat
 gttcaagctgaaaggccctcgtatgctgcaaggcagctacaacctgcatttccccgtgaaacaccagcagaaggatgatgaggttggcactt
 gccctaggagacgagaacgctgtcctcatgccagtgagctgcttcaacagaggcgttcaagaaagcaagaagccttgggactagggga
 cctggatttctcagcggtttacgaggtactgaaggcgaggtggctcagggaaggcgtga
 >29950 OsPN29950

gcttctgaggaatcaacaacagccaacttagtggggcagatcacaaccacctgtacagatgataattctgtgaacagatcagcagaaaattct
 tcacagaagaacattccattggatggagtatctgcacagtcattaagccatctgcaagtagtaggtctgaaccagtaggtcttggagggtt
 tgcagcctaagaggcggagcagaacagcaagccacctgggagtagcagtgatactggcgaagtcgtcaattcctctcgcacagcaata
 gccaaaatgctgttcaatgggccagcaggttctgcaagcccttgccttcaaaataactaatgtaaacagaagccatgttacggatttccactt
 ccatccactacttctcagtttctggtggaatgcctccgagaagacaggggtggtgaaggacaagtgattttggcagatgatattcaggtgtgct

Figure 13
 Page 4 of 5

aaacaacccagcttttgcaatctgttgcgaatgtagcagagcaaacaggcatgggtccgcaggtgattgagaaacatggtggaagagt
gtgcacagagccctgcaataatggatactatgagtaatttagtccaaaatgtggatgggtcaggaagaggtcaaggtggcattgactgtcta
gaatgatgcagcaaataatgcctgttgatccaagtcttgggtggagctggggctcgtcctgctggtacaaatagtggaacatccagattgc
agcctcggcgagtgacatgagagtggatgatgcttcagattatggaaattctcagattgatctacaccaagctcgtgaacacattgagcaac
atgactccccaggatacttcgggtcgggtcctcgaactgctgcacaggcttatgggtgaagatgagagtattgaggacatgcttgaagag
cttgtcagtgaccagaacttacagatgactacctgaaacttctgctcaacaagttcgccagaggatacagtcggcatctcaatccgggaa
ccagtcctga

acttcgaaatcattaaacctatcagccgtggagcatttggacgtgtttcttggccaagaaaagaactacaggagacctcttctataaaggta
ttaagggaaggcagatatgattcgaaaaatgctgttgagagatatattggctgaacgtgatatcttgattacagtgggaatccttttgggtcgt
ttcttctattctttaccccggaatttgaatctagtcagtgagctgaatggaggggacctgtattcttactgagaaatttagggtgctt
ggatgaagatgtgctctatataccttgcagaagttgtgctagcttggaaatttgcactctatgcacatagttcatagagatcttaacctgac
aactattgatagctcatgacgggcatataaagttgacagatttgggtgtccaaggttggctgatcaacagcacggatgatctatctggctc
gctgttagcgggtcttctgtgtatggagatgatgaacctcagatgagcgaatttgggaaatggatcaccgagcaggagacaaaagcgt
ctgctgtcggaactcctgattatttagcaccggaaatcctttggggacaggacatggtactagtgtgattgggtgtctgtgggtgttactttt
tgaattaattgttgaatacctccattcaacgcagaacacccccagacaatttgcacaacattctgaaccgcaaaatccttggccacatgtc
ccagaagagatgagttctgaagctcaggaatcattgacaaattattgacagaagacctcaccaacggctaggtgcaaatggcgcctctga
ggtaagcagcatcaatttttaagacattagctgggataccttgcaggcaaaaggctgcctttgtccctcctcagatagtgacattgaca
cgagctacttcaccagccgttattcttggaaatccatcgacgaaaacatctatgaagcatacgaatttggagattccagtataatggaagc
tagtggcagcagtagctgtgtgagcaatcacaagatgatatgggggatgagtgtagtggttactgagttgaatctagctcaaatgtcaa
ctattcttcagcaatttctcattcaagaatctatcacaactgtctctatcaattacgatttgcaccaaaggattgaaggatga
tccaccaacgaaatctgagacgtag

>27024 OsBAA85416

atgtatagtagtccaaagcattactatacttccacctggccgtcctgtacagacggctgggccaacatcacgtacaccacacttggcc
caggaagcaatcatccggccattcatgtttacgccttctgtgggcccgttacgcgtctcctgtaccatgtggcccatccgccatgtgg
cagagagacaacaccgaaacagacagacacgaatctcaacgcagcagcggccaccagcctcccgtcctgcctccgacgcggcgctg
cggccgcccgggtgacccgccctcgcgcggcggtggtccacgaccgaccagatggccgccgcgcgtcgaggtgcctcctcgtc
accggcccggcgaatccccatcccgaacccccaaaccgttaaaccttctccttctcgtcaaccgcctcacggccgcccgcgcgc
cgccgctctgcagggcggtggggaagacgacgctggtacgcgtgtgttcgaaacctccgcgaatccacccgcacatcaacattcgcg
gcttctacacccgcgaggtgagggagagtgagggaagggtcgggttcgaggtcgtcacgctcgtatggggaacccggccgctcgcctc
ctccaaggtttccagccgcgagctgttagatggcctactgttgggaggtacaaagtagatatagcttctctgaggtctctagcattgccggag
ctgcaggtcaaggacgacactgatctttcattattgatgaagtgggtaaaatggagctgttagttcggcattctcctcgtgtaatgagatt
attgaatgaacattccagtgttagccacaattctgtaccaagattgggcccagacattccggaggtgcgaggttgaggaatcatcctgga
gcagttatttacaccttaataactgtaatagggatgccatgagagaaggtgtctacaatcattgagcagctgttgcagaaaagggaatgctg
ttaggatcatgtcccggagaagatccaacatgaagaagcatgtctccttcaaaataaagctgtgcttcagtgctgttcatcttacagcat
caagccccacctctggatctgatctccatctccataagattcagtcgacatggatgaacaccggagacaaggaggagaaaacatgggtc
cgtccagcctaataagcctgttccagttggcagtttggaggaaattattatcgtcaagcattcatggcttctccttccaccagcatgtaggcaagg
ccagctggctagaatgccaaagcaagaagattaaggcatctctcccgcatctgttaaaaatttgaagagcagacccatttccacg
gtgggagaagcaacagtttgcgcgtgatcggcgccggcgccgggtgttgcggggagctgcatgctaactgcagagaggagggcac
gagaggatcctcgtactgtggtatctcttgcaacaagtgccaccatgctgcagtacctggacttctctcatgtagcacctcaagggaagt
ggtcacacaagaagcagggcggaaggatttgaagcccaaggaatagcatggagttcacctggaggtccacaaagctatggcgtcttcc
aagaggatgtccgtactcctgcaacatgaggcaacaatacccaaaagctggctcaaccacagctcatcccaatcaagaggatcatcca
tgaggacgtatccttcagaacaaatgaattcagaagagaccaagtgctattgccaggttgatgggcatggactcacctccaatgagtacaa
ccggcggtgagcttgcgcggccacacggaggagaaaaggcaagacatgataacctccacaaggccaatgccaaagagatcttct
tgaatgtgtcacaacagcatgtctccttctgcaacacaagggtccatgaagcattcgcgaagcaagctgaggtctgtgcctacgacg
acagcatggagctcttggtcagctgagcaaggcgtacgacagcagtgatgggtaaacgcagccgcgagagcaccgcaggagga
ggagctgcagaagttcaagaaggatttgcaggcgtggcaggcgagcaggatgtgggagcagtcgagagctctggaattggagagccat
ctcgacgacgatgacgtcaggtgacatcgtgccttacaggttccagcacagagggaaggataatgccggcaagaagcacacgca
ctccaatggcgacgcacactggaggaggagcaaggagagcggcaccggcaccggcacgctcctcggcagccggacgttctcct
gacgagcgcggacgccagctcagcgggtgcgcgtgtcagatttactacgaggagagaggtgctgtcgcgaagaagatagtg
atcctgaagccctgtcccagatgagcaccgacgacatcgaggagtgctgttgggggtcggcgagatggtgaagaaggagaacaacat
ggaggccttctcagggaggtgaagaagcggtcgaaggtcagctcgaaggaggatggcttccagcagagggcgccggacaggtg
ggccggcgccggcgacattccggctgacccgaagcagattgctcggagcatcgcaaccagatcaggagaccgtcaccaaggacgtg
caticggcggtgtccgggtcggagtcgacgcggctgtaccgacgcacatcccgctcaatggccagagccagatggactacatctgcaga

Figure 14
Page 2 of 11

gatgccaggaagcacctctccgacaggctgaagaatgtgctccgacgagagccggagaccgagccgccggcactgtctcaccggagaa
 gaaccgcctcggcgtcgttaacgaagagccgagggccgaagccgagggcatgaggtggcaaggaagggcaagataaggagcaaggag
 gagaagaagcacgcgacgaggttcgacgtcaggtcttcagacgtggccaccacaaggcgtcaccaactccggcgatcactccgaccc
 ggtatcggcgaggaacctcatcaggtcgttctcggcgccggtgtcggggaccacttctgaagcttctcggaggagccgagggtgctg
 accggagcacggctgagcgcaagcaggaaggctacgggagcagggccaccgccatcgtcggaggaggagaggaaagggagg
 aaggacacgttcaacatcaagggcagggtgtccaacctgagggcagaatctggggctgagagcgaagctgttcggcaagaagctccactc
 cgccgacgagtcgccattccccgatgatctccctccgacggaacgctggtcaccgccccttccgtcctcatccacccggcgtcctccag
 gagaactccaccgaggtgccggcgagcccggtcgtggtgacgtacccctcctgatgagatgagcaggggaggttaccggagccctgt
 ctgccattggaggcttcgttcagcgagcaccgatctccctgaagatggcggttagggacatgagttcagtgcttctgaaccagagcatc
 catcgtcggagcaagctcaaacgatcaagaactcgcaaaacaagttcaatccaggacgacgacgacgacacagacgagataga
 taaccccatcaaagcttacatcagagcaattctgtatcgcgggtctgtacggacagagacgaagctctgaccaactcttctcagaccgcga
 ggtgaaaccgatccccgcgtgggtgttcgaggaagtcgagtcgtcgtcgtcctcctcagccccggccaccaccgactgcgacgtcggg
 ccaccggcgtcgaccaccgcctcctcttcgacctgatcaacgagtcgtcgtcggcggtcgtccagtcgtcgacgacgtgtgcgcttca
 gcaggtggtacggcgcgccggcgaggagatccccggcggaagaggctactggacgggctgtggaacaccgtgcaggcgtgggtg
 gcgccaccgcccgcgacggactcgcccaactcgtggacgagctgatcgccgcgacatgagcatgtcgccgtggaacggcccgttcc
 gcgaggacgtcggcgcgccggcgagagatggaggcgagatactggacgagctcgtcgacgagacgctctgggacgtgctgctca
 acgtcggcgactga

>23221 Os003316a-P014

atggcgaggcgctctccttgcctcggcgtcgccatgctcctgcttggggcggtggcggtggcgagcgccttcattgatgaagcggcggtg
 ctggccgggggctcggccatggcgcccgttcgatgcaagcagggtcgtgtacatacagaagctgccggagccggagccgaagc
 caaagccaaagcctcatcctaaacccacgcaaaacctgagcccaagccagagccggagccaaaaccagtacctgagcctgagcctaa
 accggaaccaaaagccagaacaaaacctgagcctaagcctgaacctaaaccataccagagccaaaaccggagccgaagccagagcc
 aaaacctgagccggagcctaacctgagcctaagccagaacaaaaccagaacaaaagccgtaccagagccgaagccagagccaaa
 accggaaccgaagccggaacaaaaccggagccaaaaccgaagccagagccaaaaccacaccagaacaaaagcctgatccgaaac
 ctgagcctaagccacaccagagcctgagcctaagcctgaacctaaagcctgagcccaagccacaccctgagcctgaacaaaagcctgag
 cctaagcctgagccaaaagccagaacaaaagccggagccaaaacctgaacaaaaccgaagccaaagccagagccaaaagccaaagcct
 gagcccaagccataccctgagcctaagcctaagcctgaacaaaagcctgagcctaagcctgagccaaagccagaacaaaagccggagc
 caaaacctgaacaaaaccagagccaaagccagagccaaagccaaagcctgagcccaagccacaccctaagcctgagcctaagcctga
 gccaaagccagaacaaaagccagagccaaaacctgaacaaaaccagagccaaaaccagagcctgaaccgaagcctgagccaaagc
 ctgaacaaaaccagagccaaaaccataatccagagcctaaaccggatccaaaaccagaacccaaaccacaccagaacaaaagccaga
 gccaaagccagccggagccaaaaccagagccgaagcctgaacctaaaccagagcctaagcccgaacaaaaccggagcctaacc
 ataccagagccaaaagcctgaaccgaaacctaaagcctgaagcctgagccaaaacctgaagcacctccgaagaagcacaagccggcgac
 ataccggcagcgaccgaccagtga

>20115 Os011374-4328

atggtctgcgtcgccatcgagtaccggatgcgccgggggagcgtgacagggccccggcggtccgccgacgaggaaggggaagtgc
 ggtcgtcttcatctagcgatgatgatgtcacggaggacgatccggtggtagctgcaccgattgcgtcagttgcaaagcacctggagtc
 agctaatacaatgttctcttcatatggtggataattgattttatggataatctgctgggggtgaggatgttatccgtgatgcacctagcttact
 ggctttgcatagtcttctggcatttgacgtgttcttgtgtattctgtgtccttgcttgcattggtattgctgtatgttctgcctaccttgc
 attatagctattctctatgcggtttctgatcaggaaggagcgtcagaagatgatattcgccaaattccaagatacaaaattcggaacagatg
 aacctgaaaaaaaactgctgatgagacaggaccttttgggtggaataatgacagagtgttggcaccatcaacctattgagaaagtactgc
 cctgaagatgcggagtggtgcatttgcctttcgccatgatgatggtgacagagctgcgtgaacttccttgtgggcaccatttccactgtgcctg
 cattgataaatggctgcatatcaacgcaacatgccccctgtgcaagtcaataatccggaaaagtggcagtagcagtggaagtgaagaagta
 tga

>20285 OsSGT1

atggcaaccggccggcggtcggatctggagagcaaggccaaggcgcccttcgtcgacgacgacttcgagctcgccggcgagctctaca
 cgaggcaatcaggccagccccgccaccggcgagctctacggcaccgcccagggcccatatcaagctaggcaactacactgagg

Figure 14
 Page 3 of 11

ctgtagctgatgctaacaaggccattgaacttgacctatcaatgcacaaggcttatcttctgtaaaggcgctgcatgtatcactggaggagt
atcaaaactgcaaaagcagctcttgaattgggttactcgttcgcatctggtgactcaaggtttactcgcctaataagaggagtgtgatgagcgcat
tgctgaggagcttactgaagtcctgttaagaaggctgaagatggagcagctgccccctctgttcttctttgttgaggaaaaggatgatgct
gcaaacatggataatacaccaccaatggtagaagtgaagccaaaatacaggcacgacttctacaacagtgtacagaagttgtattgacaat
tttgcaaagggtgttctgtctgagaatgttgttgaatttggtaacaaatgttaagtgtgtcgttgattgaagtcctggagaggagccgtacca
tttcagcctcgtctgtttctaaatcatccctgagaaaagcagataccaagtgtatccacgaaggttgaaataagactggctaaagctgaa
cagattacatggacctcacttgattatgataaaaaaccaaaaggctgttccacaaaagataatccctccagctgaatcggcccagaggccatc
atatcttctcctcaaaatccaagaaagactgggataaactggaagctgaagttaaaaaggaggagaaaggaggagaagctgaaggcgatgc
tgcaatgaacaaattttccgtgacatctacagtgatgctgatgaagacatgcgacgagcaatgatgaaatctttgttgatctaacggctactgt
tctgtcgaccaattggaagatgttggtcgaagaaggtagagggaagccacctgatgggatggagcttaagaaatgggagtactaa
>24060 OsPN24060

atggcctccgcccgcgtcatcccgtcgtccgcgcccggcgccgcccgcgaggtgcggccgcggttarragcgctgggtgggtgggtgg
acgagcgggacggcttcatggctgcgcggaaagtgcggcgcccaacgccatcatcgactgctgctccamctycgstccg
tcggcgagcccggggagttcagacagctgcgcgcccgggtgcagcagcgccgcccactggcgcccgtgatccacatgcagcagt
tcttccccgtcggcgacgtcgcgtacgcgctccagcagggcggtkcgccgcccgcgcccggcgaccaccagcagcagggccccg
gcgctcgcgctccccgcgccccytccccgcgcccggcgccccctgttctcggctcccactcgcacatcgcacgggtgtcaccat
catcgctccgattcgggtgcgcggcgggcggcactgggtgcgacggctggtatccgataaagatggacgtgaagttcataacaaggaagagaa
aggaatgaaggaagcagagaatgtggtgaagctaaaagctcacagtggagctcttcttctcatgaaggtgaaaaaactcctaggccgc
aagctgtgctgaagggaagcagtaaaagtgttcaactcctgtggagtatacgtcaatgacattattgatggcaagacgggttaagctgtga
agggcttaaggtttatgaagggttggtaaatgagaatgagaaaaacaagatttcttcttacttaataaacaagaagcttcttctgtaggag
ggctgaagctggcgagactgtaataattgaaaaagaccaatgaagggtcatgggagggaattattcagctgggcattcctatcgttgaa
ggcctcctgaagatgactatccaagagagacaaaagtggaggctgttctgattgtcgtcatgatctgttgaccgcttgtgtcagaaggaa
attataccaacaaaaccagattattgttattgattactacaatgagggggattattctcaccctcaccatccccctccttggtatggtagaccct
tttgatcattctgcctgacagattgtgacatggtgttcggccgagttattcaggagaacgaggtgatcatagaggtcctctgaagctccttct
cgacagggtcttctgtgtgcatgggaagagtgtgatgttgtaagcgagctattcctgctgcatgtaagcagcggtatcctactaagctt
cggaagctttatcgagaaaaaagtaacctgtgaaagtgtttcacgggttactacccggtgacaccacctcctatgccttggggctcctcca
aggccgggtaacatggcgctcattctcaagccctaaacacttggatgtgccccaaacagtggtgacttccagcgccggccattggagc
gcacacattcctccatcagatggaatgcagccacttctgtagcacctgtcctcagttgtgtgagccatgccttcccatcactgttctt
gccaacitaaccaacagcttggatggcagaagctgtccaaggtctgtccacaacgactcccagttcttggcacgggggttcttcttct
cctggatcggggtcacgactgcctcatcagatgat
>23914 OsPN23914

atggaggggcgccggcgaggtgggtgtgtacgtgctcggcccgaaccaggagcacgtcggcccctacgcgctctccgagctgcgagaa
cattttgctaattgggtacatcagtgagagctcaatgctctggcgagaaggaggagtgatggaatgcatgctgctgattcctgacctacttg
cggtggtcacaaaaaggatcaacctgatgaagggaattgaagatgatttcgataaattcaaaaggaggttatagaagctgaggcagaggt
ggaagcctcgacagacaaggctgcagataacgacataaaccaagaacatggggccgatgatcctgatgaccggccagcaacccacca
gatggcgaggacgaatttactgatgatggtactgtttacaagtgggatcgtgtcctgagggcatgggttctcaagatgacctagaagg
caaaaatgacaactatgaagttgaagacatgactttgcacatgaggaagaagttttcaagcaccagatattgctggttcaaccacattagaa
gaaaacaatgttctgcggaattgaaatcaaaagaaccaccaaaaggtagaaaaagagacacataagaagcggaagtcttctgaaaagcca
gccgataaaaagggaagcttacaacctccagatagtgggttgatctcaaaagtaaacacacatgtctacgttactggttgcctgatgatgtcac
agctgaggagattgtggaggtattttccaaatgtggaattataaaggaggaccagagacaagaaaaccccgagtaaaaatctatactgaca
gagaacaggcagaaagaaggtgatgcgctagt
>24061 OsPN24061

atggccggcgccgacgtggacgtcgggacggagctccggctgggggtgcccggaggtggcgggcgccgcccaggcgccggcgccca
aggccgcgaagagggggttcgaggagaccattgacctgaagctgaagctgccaccgcccggcatggaggagggccgcccggcggaag
gcggaggcgccggccgcccgaaggccaagaggccggcgaggccgcccggcgacggcgagaagccacctgctccaaggcg
caggcagtggggtggccaccagttcggtcgttccgaggaatatcatgaccgtccagtcagtgaaggaagaagggaaggaagctga

Figure 14
Page 4 of 11

caagcagcagcagcagcccgtgccaatgccagcggcagcaacagctctgccttctggaaggcagcatggacggggcaccctacctg
cgcaaggctggacctgaagatgtacaacagctacaaggacctctcccttgccttcgagaagatgttcggcaccttcaccgcaactgggaaca
atatgaatgaggtgaatggctctgatgctgttacaacttatgaagacaaggatggtagcttgatgcttggagatgtcccggtggcagatgtt
tgtcgagtcagcaaacgcttgaggatcatgaagggtccgaagccattggcttgcaccaagagcaaggacaagtacaagaacaagag
ctga

>24063 OsAAB28535

atgaatcccagatgactacctctcaagcttctgctcatcgagactcgggcgtcgggaagtcttgcctgcttctgaggtttgcggacgatt
catacttgagagctatatcagttaccatcggttgattttaaataccgactgttgagcaagatgggaagacaataaagctgcaaatttggg
atactgctggccaagagcgatttaggaccattacaagcagctactaccgttggtgccacgggatcattgttgttatgatgtgactgacagga
gagcttcaacaatgtcaagcagtggtgaatgaaattgataggtatgctagtgaatgtgaacaagcttgggtggggaacaagtgtgact
agctgagaacagagtggttcttatgaggctggcaaggcccttgcctgatgagattggaataccattcctggagaccagtgcgaaggatgcaa
caaattgtggagaaggcattcatgacctggcaggagagataaagaacaggatggcaagccaaggccgtacaaatgcaagcaagcctgc
aaccgtgcaaattgccgaggcaacctgttgcagcaagcagctgctgcttga

>28982 OsARCNI

atgggtggttctgcagcttctattatcgaagcttgaaaaagctctgttcgagacaatttgtgacatgtcccgataaggatcgatgggctg
cttgcaagcttccgaagctggcgggaagtggaaagcaacatacttatgtgagactgaaaatgttcgttatgtctatcaaccaattgaagcact
gtatttgcctctcatcacaataagcaaaagtaacattctgaagatctggatcttgaagctactctccaagcttatacctgaatatgtccctctt
tggtgaagaggggtgtctgaaggcagcattcgagcttctgttgcctttattgaagccatttcttggaaacaaagagaacgtaactgttga
caagttaaacaatactgtgaatggagagccatgaagaaaagctgcacaagctggttatgcaaagcaaaataatgaaactaaggatgtcat
gaggaggaaagtcactgagattgagaaaagcaagactgatagagggaagcctgacaaggaggatttggatccttgagaactccaaaca
gcttcagtacatgggcattagagggtggtggtccagggggcgatcctataatttggatagactcgttcaccacaaggccaaggccg
cccattgctcctgctcctgcatctgcttactaaagttcctggtggtatgaaattaagcaaaagcacagaagacaaccagttccttgagtcttt
gaaagctgaaggagaagtcatttggaggacactcaaccaagtgcaactcaatcaaggctcatcatatcccaccaagcgatcctatcacag
tgacaattgaagagaagtcattgtcactgtttaaagggtgaggagtagcaatttgcattcaaggaaacctggctctccaaattctgaa
tgacactgatggtttattcagttacagattgagaaccaagatgtacctggccttaactcaagacaccccataatcaaaaagagttattca
acagtcacaaaattgtgggggcaaaaagatccaacaggcccttccccagtggtcaaaatgaaacaccgttgggtgaaatggagaatccagg
agttgaatgagtcactcttccattggcagtgaaactgttggccttcggtgtccggaaatgaaacctatgtgaacattgaatatgaagttcagag
atgttgaattgcacaatgttgcactctatccctctcccgacattaggagggctccaggtgttaggcagatagacggagaatggaggtatg
actcaagaaattcagtggttgagtggtccattattctgttgatcaatcaaacccagtggttcttggagttcactgtccctgcagctgatccac
aacattctcccgaatctgttgggtttctgcatcaatacattagtgattgaaggtagccgcatccgtccgctgagggaaggtagccac
caaagtttctcagaggaaatcgttacttataactaccaagtggttga

>29042 OsPN29042

atggcggggtcgtcctcctcagcacgctcctctgcacgcagtcgccatcacccgtcaccggtggcgtccggccgcctcgtggcaccg
gctgtcctcggttcgccggggccccgcggttcccgacgctcggggccgcgccacggcggctgacggcgccggcggtggccgggggac
gccgaggacgagtkggggaaggagccggcgccggaccaggcggtgccgccggcggtggccgaggcgcccggtatgtccg
gtgacgagcgaggtcgcggagctcaaggcgaagctgaaggaggcgctgtacggtacggagcgcgccctgcgcgctccagcgagac
gcgcgcggagggtggtcagctcatcacgcagctcgaggccccgaacccacgcggcgccaccgaggcgctaccctcctcaacgg
caagtggatcctcgcgtacacctattttcacagctattcccactattgggtctggtagtgttgcctcagcttgcgaaggtggaagaaatcac
aaactattgattctgagaacttcacgggtcagaattgcatcaagtttcaggaccttagcaacaacttcagtgctaccaatgccaaattgaagt
tcgaagtcccaaactgttacagatcaaatttgatgaaggcattatcgacactccacaattgactgattctattgtactacctgagaagttgagtt
atttgggcagaatattgacctgacccattgaaggcatattttcatcaattgaaaatgcagcatcctcagttgctagaacctctctgtgcaac
ctccactaaagataccaattaggactgacaatgctgagcttgggtgtcacaacctacccttgatgatgagcttagaatctccagagggtgatgg
cagcagcatcttctactgtttaaggaaggaagcacccctcctatataa

>23949 OsPN23949

acgcgtccggacttcgataaagagttgtgattggcttgcctagcaactttaagaagatgaaggcattgatcttctgaaagacaaacaagcc
ctgcagcgactcactgaggcagcagagaaagcgaagatggaactgtctacgctgtctcagacaaacattagcttgcctttcataactgctact

Figure 14
Page 5 of 11

gctgatgggcctaaacacatcgagacaactctctccagagccaaatttgaggagctatgttcggacctcattgatagccttaaaactcctgtc
 actaatgccttgagagatgccaaactgtctgttgataacctggacgaagtgttctgttggtggatccactcgtatccctcctgtgcaagagct
 tgtgaagaagatcactggcaaggatcccaatgtcacagtcacacctgatgaggtgttctcttggggcagctgtacagggtggagtttggc
 cggagatgtgaaagatgtcgttctcttgatgttactccattgtctcttggctggagacgttgggtggagtgtgaccaagattattccagaaa
 cacaacactgcccacctcaaaatcagaggtattctccacagctgcagatggacagacaagtgttgagataaatgttcttcaggaggagagaga
 gaggttgtccgggacaacaagtccttggagcttccgcttgatggaatccctccctgcaccacgtggtgttccacaaattgaagtcaagtt
 tgatatgatgccaatggtatcttctgttctgtctattgataagggtactgggaagaacaggatatcccatcaccggcgcttagtacactg
 ccaaaggatgaggttgagagaatggtggaagaggctgacaagttgtctcaggaggacaaagagaaaagagatgccattgataccaaaaa
 ccaggcagactctgtgtctaccagactgagaagcaactgaaggagctagggtgacaaggtacctgcacctgtgaaagagaagggtggatg
 caaagctcaacgagctcaaaaggccattgcgggtggatcaacacagagcatgaaggatgccatggctgctttaaagaggaagttatgc
 agatcgccaggccatgtacaaccagcagcctaattgtgtgtgtgtgacactactcctgggtgccgatgtggaccgacaagctcaggc
 ggtaagggaccgaatgatggagatgtattgatgcggatttactgacagcaattga
 >20696 OsERG3
 atgtgtcaggggacgctcaggtgtgtctgtcgtcgagaccaaggccctcgagaacaccgactacctgtgcaacatggacccgtacgcggt
 tctcaaatgccgtcgcaggagcagaagagcagcgttgcgtcaggtaaaggatctgacctgaatggaacgaaacctttatgttcagcgtc
 actcacaacgctacagagctcatcatcaagttgatggacagtacagtggtgacggatgatgattttgttggaagaacgatttcttggaa
 gcaatctatacagaaggaagcatacccccactgtttataatgtgtgaaagaagaataaccgtggagaaatcaaagtggtggcctgacgtt
 cactccagaggatgatcgcgatcgggtttatctgaggaagacattggtggatggaagcagtcatttga
 >31085 OsPN31085
 atggccgccctcttctctctctctccacgtatcagacgccccaatcaaccgggacgttccgtgtccacgacctgcgccgtc
 gctgaccaamgccgaagccgtcctcggtggggcgacccaacgccgccgacctgtgcgccgatggccgcacatctcctgcgaccg
 cgccggccgctcaacaacatcgacctcaagaacgccggcctcgccgggacgctcccttcaccttggcggtcgcgacctccagg
 acctcagcctccagaaccacaacctctccggcgacctcccgcttccgcgcatggcctccctccgccacgcttctcaacaacaactc
 ctccgctccatccccggcgacttctcagcggcctcacctccctctcgtcatctccctcgaccagaaccgctcaacgtctcctccggggg
 gtggacgatccccgccgacgtcgccggcgcgacgagctgcagagcctcagcctcaacggatgcaacctcaccggcgccatccgggac
 ttctcggcgccatgaacagcctccaggagctcaagctgcctacaacgccctctccggacccatccctccaccttcaacgcctccgggt
 gcagacgctctggctcaacaaccagcagggcgtcccaagctctccggcagctcgaacctcatgccaccatgcccaacctgaacagg
 catggctccacggcaacgacttctccggccccatcccgactccatcgccgactgcaagcgctctccgacctctgcctcaacagcaacc
 agctcgtcggcctcgtccaccggcgctcagtcctatggccggcctcaagtcgttcagctcgacaacaacaatctcttgggccccgtccc
 ggcgatcaaggcgcccaagtacacctactcacagaacgggttctgcgccgacaagccgggggtcgcctgctcgcctcaggtgatggcat
 tgcttcacttctcggcaggtcgtattacccaagaggctggtagctagtgtccgggaacaattcctgctggattggctgggaattagct
 gcgttcaggggaatgtgacgatgtcaatttgcggagatggactcaatggcacaatcagcgacagcttgggaatctgagtgaactctc
 gacatcaacctcatcggaataaccttactggcatgtgccggatagcctcacaagcctcaggtgtgcagaagctggacctgtccggga
 acgacctgaccggggcgctgcccaccttcagcccgagtgtgaaggtgaatgtgaccggcaatctcaactcaatgggacggcaccaggat
 cagcgccatccaaggatactcctgggtctcgtcgtccaggggcgccaaccttgcctggccaggggccttccggaaaataagaagaaaa
 gatcagctgttgtgtggccaccacatcccggttgacgtcagtggtgtggtcttggctcagtggtgctcattctccgtaagaaaag
 gggatcagttccacaaatgcagcttccgtcgtcgtccacccccgcgagaattctgaccagacaacttggtcaagattgtcatggtgaaca
 atgatggcaacagtagttcgacccaaggcaatacacttagcggggagcagcagccgcgctagtgtatgtcacatgatcgatacgggcaatt
 cgtgattgtgtgcaggctccctcggtggcgcaaccaagaactttaccaggacaatgtgcttggccgtggggggttgggtgtcgtctacaagg
 gtgagctgcatgacggcacgatgattgtgtgaagaggatggaggctgcagtgatcagcaacaaggccttgatgagttccaagccgaga
 ttaccattctaaccaaggtacggcaccgcaacctggtctcgatactgggttactcaattgaagggaatgagaggctgctgtctatgagtaca
 tgtcaaacggggctttagcaagcatctgttcagtggaagcagtttgaattagaacctctcatggaagaagaggctcaatattgcactgg
 atgtgtcgtgtggatggaatatctgcacaacctggcacatcagtgctacatccatagggatctcaagtcagcaaacattctcctcggcgacg
 attccgagcaaaaggtatcagactttgggttagttaaacatgcaccagatggaaacttctgttggaaccagactcgtggaacttttgggta
 ctggctcctgaatatgtgtgactgggaaaatcacgacaaaagcagatgtcttcagcttggggtagtgctgatggagctgataactggaat
 gactgccattgatgagagccgtctggaggaggaaacccgttatctggcctcatggttctgcagataaggaaggatgaagataggctcaga

Figure 14
 Page 6 of 11

atggggagggtgacgacgaaggcgaggcggaaggcgaagagcatccccaccacggcagctgcacctctcctgtacacgaacgac
atgaccagcgacacggcgacgctgccgacgagcagcagagggaaccaccactgcggcacctgctgtaccgccgcatcagcttcgct
gcacgtccatcttcgccggcgaggctcatcgccgacctgctctgctccagatgtccccgccgtggaacagcgccacgtgcacgacg
tcgccgtgaacctcaggaacccggacccgatggagatggcgaacagcgggctcaggtacagcttccgtagctctgtacgccgccacg
ttgaggtcgaacctcctctgtgagcaccttggtggtgtcgtacttctgccggacgccgtgaacagctggttggaagattgggaacttctc
gcgtcgaacgtgtcgaacttccagtagcagagcggcacgatgaggtagatgaacatgacgaagcccgcgccggtgttgccgatggacg
accacggcgccaccagcgggctcccgtgtacgccgagatgccggccagtcaggtgtaaggcgccgacgccgaggccgtgtagc
cggag

>29984 OsPN29984

gcgccgtcaagctcatctggaaccacatcaaggccaacggcctccagaaccagcaaaaaaggagattaactgtgacgataagctta
aaagcttattcgccgggaaggacaaggtcgggatga

>29983 OsPN29983

ggtgtggagacggtgtgacggacacggcggggaacaagggtggtggtcaccggcgccggcgagctgaaggagcgcat
cgaggcgccggaccaagaaggccgtgcagatcgtctccgccggcgccggccgccccaagaaggataaggaggagaagaaggac
aaggacaagaaggcgccggcgacgacaagaaggccgagaaggagaaggcgccggcgccggcgataagaaggcggaaggga
gaaggcgccggcgacaagcccaaggaggagaagaaaggcaaggagcccaagaggagacggtgacgtcaagatccgctccac
tgcgagggatgcatcgaccgcatcaagcgccgcatctgcaagatcaaaagggggca

>30844 OsPN30844

gtcgccgtgcgaagaaggagacgcgcgacggcgccatgcagaccatgccgtcgcgcgtgcagcagggcgccgccgacttcgagga
gacgtctctgtgcgtgccaccttactgcagcggcgccggcgaccggcaagccgtcaggttcgagccccggcgcttctcctgtc
ggcggtggcggtggaagccccgagctcgaacttggccggagcgccgtcgaacctgagcctcctctgaaggagtcacggacaagagc
cagcaaggggagcgtgtccggcaatgggacatggcggtgccgctcggcggaaggcggaaggcgcgagctcgtcgtcaaaactgtcgt
tccagatcatggacgacggcgcgctcggggtgttcaaccagaccggagcagcaaccaagattaactcgtcgtcgtcttcttcttctg
cacggaagcagagcaagctatccttcagatcacgagccggaaggtgtcgcggctcgagccgaagctgacgccgacaaagggtcggc
gtcgccggacttgcgaggcattgacgacttcaagctcagcagccagtttgccatcgctggcagaggccaagcaagagcagaaggagc
cagagccggcgagccggaggagaaggctgatgactcggagttccggagttcgacgtagtgacaaagg

>30868 OsPN30868

ctggacgaggtgatggcggtgagccccgtggggctaggcgggcggtcgcggcagatattcgacgaggtgtggcgcaagtctcgcggc
tggggcagatgtcgagcgcgtcgtcgacggcgctggcgaggagagcagggcgtgctcatccggcgccggacccatgtgcgagttcac
cgtccccggcgcgacgacaccaccgtctcgtcgtcggcgccaccagccgcatcgccgcatcgtcgtccgcaagctcatgctccgcg
ggtacaatgtcaaggctttagtagaagggaatgatgctgaagtatagatatgtccccagggtcggtggatattgttcggtgatgttggtgat
ccttcaactgtcaaatcggtgtatcaggttgacgaagataatctattcgcaactgcacgtcaactattactggagacctaaatagggttg
ataatcaaggagtaagggaatgttagtaaggcttccaggattactacaatgaattggtcagcttagggctggtaaaagcagcaagagcaaa
ctctgtatgcaaaaatttaaatctcccaagtctcgaatggtgggaggtggaccaggatcttattttcaaacacatttgccttctaggttgatg
aaggatcgtatgcatatttgacttttcagaggctggtcagctgttttctcaggattgttttcacaagaggtggatattgtgagatatcaaaaag
gctatcttcttcttgggactactctagacaggtatgatggcttacttttctgggagggaatggaagatcatatgttatttcttgagactg
gtccgctgggtgacacctccagagcaagaagtacttgcgtcgatgactacaaaagtaggattttgtaggggtgagggtagcattttcagcttt
tcggccagtaaacctcaagatcctcccctagaccttttctgtgcatactaacaattagggttgaacccaaaagacagagacctggtgat
ggatctcaaaagtgtactgaccccagaaaattttgagcttatattggaatacatcaaaagcittgcctactggtcaagaacagacttattctggtc
tcatgtcaggttctgggattgaacctaacagaagagaacaagtcctcaaaagcaaaaaaggctggagaggatgattgagaaggctcaggc
cttgatatacaaatagtagccctggtccactgcaggagaacctggtggtcagcgtgattgatctttgatcaagggaacagaatctcgca
gggtatcagttgcgccgatgtagcagatattgtgcaaaagcattgcatgattccactgtaggaacaaaagtttgatgtatgctacgagtatg
tagctaagcaagggaatgagctatatgaacttgtggccatttaccagataaaagctaacaactatcttacaccggcattatcttctagagaa
gaacacctga

>24292 OsBAA78745

atgatacggggcgtacgcgcgtgcaagacggcggcggaggagcgggcgggtggtgcggcgggagtgccggcggtacgatacgggcggcca
 tcagcgagggggaccaggactaccgccaccggaacatggccaagctcatgttcatccacatgctcgggtacccccacccttcggccag
 atggagtgccctcaagctcatcgccggcggggttccccgagaagcgcacggttacctcggcctcatgctgctgacgagcggca
 ggagaatgctatgttctgcttgagcatttggtagtacttaacgcggacagagatcttaaccactcgaaccagttcatttggggccttgactct
 gtgctcttgcaacatttgcctcgctgaaatggcgcgtgatctgtcacctgaggtggagaggctattgcaaagtagggaaacaaataccaag
 aagaaggctgcctatgctctataaggatcgtacggaaggttcagatttggcagagaacttcatgggctctgctgttctactactgaaggaaa
 aacatcacgggggttctcatatctgctgttcagctctgcgcagaacttgtaaagcaagcaaaaggagcattggagtacctgaggagaactgc
 ctgatggttggcagaactactgagagatgtgtccaatagttcatatgctcctgaatatgacattgctggaattaccggtcgttctgcatatca
 gagtgcctaaagctcatgcgaatttgggtcaaggagatgcagattgcagtgagtttgaatgatattctgctcagctgtgctctgtatcttatat
 cgaattattgcatgcatccaggttgaacaaaaactgagtcataaagaacgcaggaaatgctattttatatgaatgtgtgagactataat
 gggcatcgaagctactagtgtttacgtgtgctggcaatcaatatttgggtagatttctgtccaaccgtgataataacatcagatatgtgctct
 gaacatgcttatgaaggccatggaggtagacacgcaagcagtgagaggcatagagcaacaatattagagtgtgtcaaggatgctgatga
 tctattcgcaaaaggcccttgaacttgtttaccttctgtcaacgatgcaaatgcaaaatctttagcaaggagctgtgttattacctggaagta
 agtgatcaggactcaaggacgacctcacagcaaatgatgtcaattgttgaagggtttcccaagataaaacttggtagcttagaccagatgtt
 caagggtttatctctggctggaaattatgtgaaggacgatgtatggcatgctctaatagtcttaataagcaatgcacatgaaactcaaggatactc
 agtgatcattatacaaggcattgctagcttgcgggtgaacaggaaagtttggtaggtagctgtatgggtcattggtgagtagtgtaaatg
 ctggtgaacaatgttggatgctggacatagaggaaacacacggtaacagaatctgatccgtggatgctgtagaggtctctttaaagca
 tactctgcagacgtgacaactcgggctatgtgtctagtatctcttgaagctcttccccgattccaccgacttcagaatcatcttctgagc
 ggatgcctgtgatagatgaagctagttagtggtaagagagctgttccacacaagcaactatttcatcagataaattagctgtcgcagcaac
 tcttgaagctcgttaagcttccaaatgggttagcaagccaccaccggctccttagctgatttgccttgaatttgcagatgctcctg
 cgactacttccgcccctactacagcacctaatgatttctacaggatctttagggcataggcttgcctgatacatctacagcaggtttgatctta
 ctttttgcattcagtagacataatttgggttgacaagtgaagcaagtgtgctcatatttccacatttgcactcatttatgtggagctccatcagcaa
 gcacagatatctgatggatcttctatcaattgttcttccagtagcaaaatggcccaccaacagtatcaaaactttagccttctggtaagatg
 agaatacagcttaccgccaatcacagcttccagagtgcagcttgaagatcatttcaatttgaagcagctctggaaaacctcaggagaga
 ctacaattcatgtagctttacaatttgacatcctaatactacggatttcatctttaggcagctgtaccaaagttatccagttgcgttggac
 cccgtagcagcaaacacgcttctgcccagtggaaatgattctgttacacaaagcctcagtgacacaaataaccaacatggacagaaacccct
 tgcgatgcgtatccggataacttacaagtgaaacgacaaactcatcatgaatcggaggagatgctcgacaaaatatgacctggtaaagcact
 cgagctccttggccatcgaggcgataaacctccctgtaccgaagtgaacgagcactagtgcacgagtagcttagcctgtgtctatcgttct
 ggccgttgcgaacgctcccgtccaagcaataagacgagacgtggctgagatataa
 >30845 OsPN30845

ggcgaccaggggaggacgtacctgttcagggtgtccaatgtcggggtgaagacgtccgtcaatgtcaggatccaggggacactcgtgag
 gttggtggaggtggaggggacgcacccgggtgcagaacgtgtacgactcgtcgcacgtccatgtcggccagtcggtggcgttctcgtcac
 gctcgacaaggcggc

>29997 OsPN29997

atggggctgctcaccagggcgaggaggaggagacggcggcgccgaggagtggtcggcgaggcggtcgtctacgtcaacggcgt
 ccggcgctgctccccgacggcctcgtcacctcacctgtccaataacctcagagacattggtcttctggaacaaagcttggatgtggtg
 aagggtggctgtggagcctgcactgtgatggctcatgctatgatcaactacaaagaagacacagcatttgaatcaacgcattggtgctc
 cgctttattctgtggaaggatgcacataatcacagtagagggaattgggaatcgtcagcgagggttgcacccaatccaggaacgtttagcc
 atggcccacgggticacaatgtggatttgcacccctggtttgtgatgtcgtgatgtgattgctgagatcaagtgaacagcctcctactgaaga
 gcagattgaagatagccttgcaggaaattatgtcgtgtactggctacagaccaataatagacgccttccgtgtttctgaaaagagatgat
 cttttgtacaacaattcatcttgaaaaatgcagatggccgacctatctgcccctcaacaggaaaacctgttctgaggatcagaaagac
 atcaatgtagtgaatcttatttgcacacctcagaaaagctactcaccttgtcatacaatgagattgatggaatgcgtacagtgaagaaga
 actcattttcccccgaaactcagttgagaaaagtacgtcacttaaatgaatgggttaattgggttgcgtggtatagacctttaaactaaa
 gcaagtattgcatttgaagcatgtacccaaatgcaaaactaatcatttggtaactctgaagtaggagttgaaacaaagttcaagaatgccca
 gtataaggcttgcctcagttactcatgttccagagcttcatacccttaaaagtgaagaggatggtatataattgttcttctgtaagacttga
 cagctccaaaatttctcagaaagggtatttttagagcgtgattcacatgaatttcatcctgtgaggcaataactcggcaattaaatggttgc

Figure 14
 Page 9 of 11

gggacacaaatcaggaatgttgccttctgttgaggtaacattgtactgctagccaatatcagatctaataccactttggatggctacaggtgc
 aacgtttgagataattgatgtgaacaataatattaggaccattctgcaaaagattttttctgggtatcgtaaagttgacttaaacctgatgag
 atattgctctctgttatactgcatggacaaggccatttgaattgtcaaagaattcaagcaggcacatagaagggaggatgacattgcgttg
 tgaatgctggaatgcgtgtctatataagaaaagttgaaggtgattggataatttcggatgtttcaattatctatggaggggtagctgcagttcac
 accgtgttcaaaaactgaacaccttttctactggaaaaaaatgggactatggattgttgataagacttttgatctgttgaagaagacgtagtt
 ctggctgaaaaatgcacctgggtggaatggttgaatttcgcagttcacttactttgagtttcttttcaaatttttccctcatgttactcatgaaatgaata
 taaaaggattttggaaggatggattacatgcaaccaatctttcagccatacagcttttactagacctgtcgggtgttggaactcaatgttatgaat
 tggtagacaaggaactgcagttggccaacctgtggttcacacatcagctatgcttcaggttactggtgaagcggaataactgatgacacac
 cgacaccccccaatacctgcatgctgctgtggtgctgagtagcaaaagctcacgcacgcatattatctattgatgcttcacttgccaaatcttcc
 cctgggttgcgggtcttcttcttcaaaagacgtgcttggcgtaaccatactgggctgttatccatgacgaggaggttttgcattctgatgtt
 gttacatgtgttgccagattgttgacttgttggcagatacccgatgaatgcaaaagctgctgcaataaagcaatattgagtattctgaa
 ctccagcaattttatccatagaggaagctgtgaaagctggtgactttcatccaatagcaagagatgcctagtaaaaggaatgttgaacaat
 gttttctgtcgggtgatgcatgataagaattatagaaggaaaagtacaagttggaggtcaagagcacttctacatggagcctcagagcactcttg
 tatggccagttgattctggaaatgaaattcatatgatttcatctaccaggtccccagaagcaccaaaagtagttgtaattgttcttggcttcc
 acaatcaagagttgtttgcaagactaagcgtattgttggtggttgggtggaagaaaccagatcagcaatatttgcagcagcatctgta
 gctgcttattgttaaggcagcctgtaaagctgttttggacaggatattgacatgatgacaactggacagaggcacagtttctaggggaagt
 acaaggtgggattaccgatgatgggaagatattgacctagacctgtgtttataacaatggtggtcattcacatgattgtcccttccagttcc
 tggagcgtgctatgtttcattcagacaatgtctatgataacaaatgtcagagtcaatgggcaagtattgttcacaaatttccaagcaatactg
 ctttcagaggttttgggtccacaagctatgctgattgcagagaattggattcagcacatggctacagaactcaagcgagtcctgaggaga
 taaaagaacttaattttcaagtggaggtatgtgttcttattgcccaggttgcctcaaaattgtacaatacattcagtagggatgaactaaagg
 tttctgtaatttatggaagctcgaaagctgtaattgattttaacaataataaccgttggagaaagcgtggcattgctatggttccaccaagttt
 gggatacttctactacaaaattcatgaatcaggttctgaggatgaacaattgcaggctgtgcttagtgcaagtttactgatggaactgtcct
 tgaacgcatggtgggttgaaatggggcagggtttacacacaaaggtagcccaagttgctggcttcatcattcaatatccctcttagctctata
 ttatctcagaacaagcactigataaggtaccaaatgcaacaccaacagcagcctctgctagttcagatttataggtgtgcagtttggatgc
 ttgtcagcaaatattggctcggatggaacctgttgcctcaagaggaaaccacaagtccttgcaggttgggtctagcatgctacctggaaagg
 atagatctctctgctcatgattttatcactcctgatgttgggttactgggtgtctggcaagggaactccatttactattcacatcaggag
 cagcatttgcagaagtgaattgataccctaactggggatttccacacaaggacagtagatattgttatggtcttggctgttcaattaatccg
 gctattgatattggccagattgaaggagggtttatccaaggattaggttggcgccctgggaagaactaaaatggggggatgataaccacaa
 gtggattgcacctggacatctttcacttgtggcctggcttcaaaaatacccttctgtaaatgatataccttcaaaactcaaggtctcactttga
 agggcgttttgaatccaaaggtcattcactcaaaaggtgtaggagagccaccgttttctcgggttcagccgtctgttggccataaaggat
 gcgatactgccgaagagctgaggagggtcacttcgactgttcccactcgcagccagcaacaccggaaagaataagaatggcatg
 cgtggattccatcacaaagaaatttctagcgtatattaccgtcccaagcttagtgtag
 >30843 OsPN30843

ggtgctggcctgcaaaatcttgggaacacttgctacctgaactcgggtgctgcaatgcttgacgtacacagaaccatttgcggcctatttgcag
 agcgggaagcacaagtcttcatgtctactgctggattttgtgcaactatgtgctcttcaaaaccatgttaagactgctctacagtaactggaaa
 aatagtgcgccatcacagattgtcaagaacttgcgtgcatctcccgtagtttccgcaattcaaggcaggaggatgcacatgagttaatggt
 taatttacttgaatcctatgataaatgttgcctaccttctggagtaccaagtgtctccttagtgcataatgagaagagcttagttcacaaaatattt
 ggtggccgcctaagaagtcaggtgaaatgcacacagtgctcacattgtctcaacaaatttgatccttcttggatctcagccttgatattggtaa
 ggccacgtctctggtgagggcacttcaaaatttactgcggaagagctcttagatgggggagaaaagcaatatcagtgccaacgctgcagg
 aaaaagggttagcaaaagaaaagttacaattgataaggcgccatattgtctgacaattcatctgaagcgcttagcccttcaacccccgtg
 aaaagattgacaaaaaggttatttgcagcaatgcttagactgaagccatttattagcagcttaaaagtgaatgtgaatacagcccttatg
 gtgttttgggtcatgctggttgaacactcagtcaggccattatttcttctgctgaacttccagtggatgtggcacaaccttgatgataacc
 aggtacgccaaagtcctgaggcagatgtattgagacagaaagcatatgttatatttattgtacgtgacagagttgggaatccgactccacgca
 aggataatatcactgctaatatgccagccaggagaacaataacctgaaaagatctcaggtctgagtgtatgatccagagcggcgtaattgaa
 gcaaaattgaacggttctcgtctccttattggcgataagaggttgacggcataagcaatgggaactcaatcaagacttcaagagagcattat
 ttgaagaaagatggcaaaactgaagctcctaagccttctgaaaacaatggcctggctccacacagaagcttctgcacctcaaatgatgg

Figure 14
Page 10 of 11

tgctaccttatctgctcaatctaagcaaatcttcaactgggtcacagagaggtatcttcatcagatcgaatcgcattcttacacatgtgattgtg
 aaccaagcagtggtccatggtccgtcgcaggaattgcaaccaagggtgatggcttaaccgatacctctccctgggcaatggcaatgcgat
 attgtctgagcgaaataaacaacatcccagcatcaaaatccatttctatgccagctcccatggaaaggacacaggggcagggcttgctg
 cacaactttcccaactaaggatgctattgtttcaaatgggtgtgtacctagtagcagggatccaattccagtgagaagggtgtgtggcttgcaa
 aatccatcaagcaagatgacaaaacgggtgaaagaacttctataagtgagaacaatatagtatcaggactgaacgagtcaacgccagaa
 aacagactagctctgaagttccatgaaggtagctgcagctgattcgtgcaatagcaacacacctaaggggtgattgagtgcaagaaa
 cttgtgagatatccagtcataacatgtggcttgggccagacaagttatgctagggtcactaaagggtacaaaagaaaaagaaatgcaatag
 aactagaagacggctgtagttgtgaggacatggccaatgccacttgttcaggtacaacacaagcgagcagcagggcatcaacatcaaca
 acaacgtcgtctgaaactgtgcaatgtactcccagagggcggaacgtgcctatgatagtacagtcctaaaaataataaccagaagcaaaa
 acaagcaagacgttattggagccgatactggtagtgtgaacttaacatggataagagaaatgtattagtgaaccgtgctagtgtgagc
 ttccgaagctgggtccaggctcttctgcaaatcaggaacattcaaggaacaatgttcagtcataaattgggaggtcccgccatttactgttctg
 acaagggacttggctgaagttactgttccatgctgggatgatgtgtgtgtgcaaatgctgaggcaagagagtgcaaacattcagaaagcaa
 gagcattggatatgtgttagatgaatgggacgaggagtgataccgtggaagaccaagaaaatcagaaattcaaaggaagactacgggtg
 accaaacccttccaggaggaggccaactacatctcacagagaaacatgaaacagaggacatatcaaccaaatcctggaagaacatg
 cccatgttagaagatga

>30857 OsPN30857

caggacacccgtcctctgcaagcccaacgtcgacgccatggaggaggccctcaggatcgccaacgtgaaccctcacaaggcgatttctt
 cgacgacagcgtgaggaacatccaggcggggaagaggatcgggctccacagggtgctggtgggcacgccgcagcgggtgaaggggcg
 cggaccacgcgtggagagcatccacaacatcaggggaggcgctccggagctgtgggaggaggccgagaaggcggaggacgtgctc
 atctactccgaccgcgtgccatcgagacatcggtagccgcgtaggccacacactactgtctctccgactccggcgaatcgctcagaacga
 gcaaccatccatccatcatgtccatccctggccgggcagggtatcatcagggtgggaattcagaattggccgggacaacaatacgttcaa
 gaaaaaagaggagggtgttccgtggaagttcaggattcagaatgtaa

Figure 14
Page 11 of 11

Figure 15

>19651 OsCHIB1

atgggtgaacggctacttgttccgggagtagacatcggcgcgacgttaccggcggtgcgttctccgacgtgcccgtcaacccgggacctcagct
tccacttcafcctcgccttcgccatcgactacttcatggcgacgcagtcctccaagccggcgccagccaacggcggttccgccccgtactgg
gacacggccaacctgtccccggccggtcgccggcgccaagggcgccgacccccaaacctcagcgtcatctcgcctcggcgggcgac
accgtccagaacaccggcgtaacgccacgttcgcggcgacgtcgtccgtcgacgcgtgggtgcgaacggcgccgactccgtctccg
gcctcatcgacgcctacggcctcgacggcggtcgacgtcgactacgagcacttcgcccggcggtggacacgttcgtggagtgcatcggtc
gcctcctcaccgagctcaaggcgcggaacccgaacatcgccacctccatcgccggttcgagcaccgggtgggagcgctactaccag
ccgctgtggcggtcgactacggcggtgatcgactacgtcaacttccagttctacgggtacggcgccaacaccgacgtggcgacgtacgtg
atgttctacgacgagcaggcggggaactacccggcgagcaagctgctcgccagcttcaagaccgggaacgtcaccggggtgctctcgcc
ggagcaggggatcgccggcggaaggagttgagcgggcaggggaagctgccgggggtgttcatctggtcagcggtatagctccat
ggtcagcagctacaagtttagtagtagaccaaggctcaggagatcgctcgccaaccactga

>19707 OsCS

atggcgggcgaacgcgggggatgggtggcgggatcccgaaccgggaacgagttcgatgatccgccccgacggcgacgcgccaccggcg
gctaagccaggggaagagtgtaaltgtcaggtctgccagatttggcgacactgttggcggtcgcggccaccggcgacgtctttgttcctg
caatgagtgcgcttcccgggtcgccgcccgttctacgagtagcagcgcaagggaagggaaccagtgctgccccagtgcaagactagata
caaggagcacaaggtagccctagagttcagggcgatgaggaggaggaagatgttgatgacctggacaatgaattcaattataagcatgg
caatggcaagggtccagagtgagcagatacagagacaggggggaagatgttacctgtcttcatcttctcgccacgaacaacatcggttccc
cgtctgacaagtggggaacagatctcaggagagatccctgatgcttccccgatcgccattctatccgcagcggaacatcaagctatgtga
tccaagtgttccagttcctgtgaggattgtgaccctccaaggactgaattctatgggattaacagtggtgactggcaagaagagttgcc
agctggaggaacaagcaggacaaaaatgatgcaggttagctaataatattccagaggcaagagggggagacatggaagggactgggt
caaatggtgaagataccaatggtgatgatgcacgtctacctctgagccgcatagtgccctatcccttcaaacagctcaacctttaccggat
tgttatcttctccgttctatcctgatgttcttctccaatctgtgtcactcatccagtgcgggatgcttatggattgtggctagatctgttatct
gtgaaatttgggttgccttacctggtccttagatcaattcccaagtggtacccgataaaccgtgaacataccttgacaggcttgcatgaga
tatgataggaggaggagagccatcacagcttgcctccattgatgtcttgcagtacggtggatccactaaaggaaacctcctctgatcacagca
aacactgtttgtccattctgggtgtggattaccctgttgacaaagtgtcatgctatgttctgacgatggttcagctatgtaactttgaggctctg
tcagaaactgcagaatttctaggaagtgggttccgtttgcaagaagcacaatattgaaccacgagciccagagtttactttgctcaaaaaat
agattacctgaaggacaaaaatccaaccttcttgttaaagaaaggcgggcaatgaagagagagtagaagaattcaagggtacggatcaatg
ctcttgttgcgaaggcacaataatgaacctgaagggggtggaccatgggtgatggcactgctggcctgggaataaccaagggtacc
ctggcatgattcagggtgttcttggggcacagtggtgggttgacactgatgttaacgagttgccacggcttgcctacgtctctctgaaaaga
ggccaggattccagcatcacaagaaggctggtgcaatgaatgcaattgattcgtgtatctgctgtgacaaaatggtgcctatcttctcaatgt
ggattgtgatcattacttcaacaatagcaaaagctcttagagaagcaatgtgcttcatgatggatcccgactaggaaga

>20775 OsHSP70

atggcgggcgaaggcgaggggtccggccatcggcacgttggcacgacactactcgtgcgtgggcgttggcagcacgaccgcgtgg
agatcatcgccaacgaccaggggcaaccgaccacccctctacgtcggctcaccgactccgagaggctcatcgagatgctgccaag
aaccagggtcgccatgaacccatcaacaccgtcttcgatgccaagcgctcattggcaggaggttagcgatgcttctgttcagagtacatt
aagctctggcccttcaagggtgattgctggacctggtgacaagcctatgattgttccagtagaagggtgaggagaagcagtttctgccga
agagatctcctccatggtcctcatcaagatgcgtgagattgctgaggcctaccttggcaccaccatcaagaatgccgttgcactgttctgcc
tacttcaatgactcccagaggcaggccaccaaggatgctggagtgttgcgtctcaatgtcatgcgcatcatcaacgagccaactgtgc
cgctattgcctatggtcttgacaagaaggccaccagcgttgggtgagaagaatgtcctcatcttgaccttggaggtggtacctttgatgttccc
tcttaccattgaggagggtatcttggaggtcaaggccacagctggtgacacccatcttgggtggaagatttgaacaccgatggtcaacc
actttgtgcaagaattcaagaggaagaacaagaaggatatactggcaacccagggtctcaggagggtgaggacagcttgtgagaggg
cgaagaggacctgtcctccactgccagaccaccattgagatcgattcctgtatgagggcacgatttactcaaacatcacccgtgcc
ggtttggaggagctcaacatggatctcttcaggaagtgtatggagcctgtggagaagtgcctcagggtatgtaagtggacaagagctctgtt
catgatgttcttctgttgggtggtcctactaggtacccgaggggtgcagcagctcctgaggatttctcaacggcaaggagcttgaagaac

Figure 15
Page 1 of 6

atcaaccagatgaggctgttgccttatggctgctgtccaggctgccatcttgagtggtgagggcaacgagaagggtccaggacctctcct
gttgatgttacctcttctctcgtttggagactgctggtggtgcatgaccgttctgatcccaagggaacaccaccattcccaccaagaag
gagcaggtcttctccacctactctgacaaccagcctggtgtgctatccagggttatgagggtagagaggaccaggacacgtgacaacaacc
tgctgggcaagtttgagctctctggaatccctcctgctcccaggggtgtccacagatcactgttctcgcacattgatgccaatggtatcctg
aacgtgtctgctgaggacaagaccaccgggcagaagaacaagatcacatcaccaacgacaagggcaggctcagaaggaggagatt
gagaagatggtccagggggccgagaagtacaagtcagaggatgaggagcacaagaagaagggtggagtccaagaacgcgtggagaa
ctacgcctacaacatgcgcaaccacatcaaggatgagaagatcgctcgaagctcccggcagcggacaagaagaagatcgaggatgcc
atcgaccaggccatccagtggctggacggcaaccagctcgtgaggctgatgagttcgatgacaagatgaaggagctggagggcaictg
caaccccatcatcgccaagatgtaccagggcgctggcgaggacatggccggcgcatggacgaggacgatgtccccgggtggcg
cagcgggtgctggcccaagatcgaggaggtcgactaa

>20899 OsCATA

atggatccttgcaagttccggcgtcgagctgttcgacacgaagacgacgacgaacgcgggagctccggtgtggaacgacaacga
ggcgctgacagtggggcccaggggggccgatcctctcagaggactaccctgatcgagaagggtggcgacitcgccgggagcgcac
ccggagcgcgtgtccacgcccgcggcgctccgccaagggttcttcgagtgcacccacgacgtcaccgacatcacctgcgccgactt
cctccggtccccgggcggccagacccccgtatcgctcgttctccaccgtatccacgagcgcggcagccggagacgatccgcgacc
cgcgcggttcgccgtcaagttctacacccgcgagggcaactgggacctctcggcaacaactccccgtctctcatccgcgacggcat
caagttccccgacgtatccacgccttaagcccaaccgcgctcccatgtccaggagtactggaggggtcttcgacttctgtcccaccacc
ccgagagcctccacaccttcttctctcttcgacgacgtcggcatccccaccgattaccgccacatggacggcttcggcgtaaacactac
accttcgtcaccgcgacgccaaggccaggtacgtcaagttccactggaagcccacctgcggcgctcagctgcttgatggacgacgaggc
cacgctcgtcggcggaagaaccacagccacgccaccaggacctctacgactccatcgccggcggaactccccgagtgaagctgt
tcgtccagggtatcgacccggaggaggaggagaggttcgacttcgacccgctggatgacaccaagacatggccggaggacgaggtgcc
gtcctccggcccgtggggcgctcgttctcaaccgcaacgtcgacaacttctcaacgagaacgagcagctggcggttcggcgccggggtgt
ggtgccggggtatctactaccgacgacaagatgtcgtcagtgagggtgttcggtacgccgacacgcagcgtacaggctggggccaa
actacctgatgtcgggtgaacgcgcccaggtgcgccaccacaacaaccactacgacggcgccatgaacttcacgacggggagcag
gaggtggactactaccatcgccacgcgcccgtccggcaccgcgcccggacgcccacacgcccgcgccccgtggtggggaggagg
cagaaggcgacgatacacaagcagaacgacttaagcagccccggggagaggtacaggtcgtggcgccggatagacaggagaggttc
atcccccttcgcccgcgagtcgcgacccccaaaggtctccctgagctccgcgcatctgggtcaactacctctcccagtgtgatgagctgtt
gggggtgaagattgcgaataggctcaacgtgaagccaagcatgtga

>22020 OsPN22020

ccacgcgtccggacggtggcctcggcgcgacctaagaacgtcaactggcgcaacggcgacctcaagccctacgccgtcgtgtgga
tcgacgacggcgccaagtgtcaccacgcgtcgacctcgacaacgccgacaacccccacctgggatgacaagctcaccgtccctctccc
ccctccacccgctcgcgacgacgcccgtctctacctcgacgtcgtccacgccaacgccaccgacggcgtaagcccctcgtcggctccgc
gcgctccccgtccgcgacgtctcgcgacaccggcataggcgcccagcctcccgtcctccgctcaagcggccctccggccgc
ccccatggacgctcgaagtccgcgtcggcgtccggagcccaagcgtactacgacccctcgcggcgtaacctgtccctaccacca
gcagacgaaccgtgacctacgcctacggtaaacactacaactgggtggtatggctatggcgtatggcggtaccccttctctataa
>22154 OsPN22154

atggcgggcgaggagcagagcgcggagcagttcagggggccaggcgaggctcccgggggtcggcgccccggcggtacgacctgc
ggctcgtccccgacctcgacgggtgcgcttcacgggatctgtggacgtctccgtcgacgtcacggcgccaccagggttctcgtgtcaa
cgccggcgagctcagggtctccccggggcggtccagttcaagccccatggcgccgagcaggagctgcaccagcggaggttaccat
gtccagaggatgagatcctgattatccgcttcaatgaggttcttctgtaggggaggggaactttagtaattgcattcaagggaactctgaac
acaagatgatggcttctacagaagtgtgtatgaactcaatggggagaagaagaacatggcggtacccagttgaacctgctgacgcaag
gctgttcttctgttgggatgagccctctttaaggctatattcaaaattactctagaagtccgtctgagaccgttgattgtcaaacatgcc
gttggtgaagagaagggtcaatggccttatcaagctgtctattttcaagaacaccaataatgtcaacatacctagtggctgttattgttgcat
gttcgactatgtgaagcttaccactgatggcactagggttcgtgtttacactcaagttggcaagagtgccagggaaggttgactaga
ggttgcgtgaagacactggctcttcaaggagtattttgctgtgccataccactcccaaaaatggacatgattgccattctgatttgcctc
tggagcaatggagaactatggcttggttacatatcgtgaacacgcttgttatttgatgagaacactctgcagctgcgaataagcaaagggtt

Figure 15
Page 2 of 6

gcagttggttagcccatgaattagctcaccagtggtttgaaatcttgtagcagtgagtggtggacacacctatggcctaacgagggtttg
cgacatgggtctctacttagctgcagacaatttcttctgaatggaatgtttggaccaatttcttgaggaatctacaacaggttcaagttag
atgccctgcagggtcacatccaattgaggtcgacgtaaatcacgttgatgaaatagatgaaatatttgatgctataagctacaggaaaggag
ctgctgtattcggatgctgcaaagctatcttggggctgaaacttttcagaaatctctggccgcatacatgaaaagttcgcatattcaaatgcc
aagactgaggacttggggctgcccttgaagaggggtctggtgaaccagtgaacacattaatgcattcatggacaaagcagcaagggtacc
ctgttgtaatgtgaagctcaaggatggaaagctagagatggagcagacacagttcctatcaagcggggctgaaggagttgggcaatgggt
ggttcccatcacactgtgtgctgttctgactcacgtcaagagaagttccttttaattgaaagcaagaggatttaattgtcagggctttaga
atgtcaaaagaaagaagacttctggattaacttaatgtcaatcagactggcttctatagagttagctatgatgaggaacttgcattctgactta
galatgcaattgaagccaacaaactgagtgacgacagatatggtgtacttgatgacacatatgccctctgtatggctggcaaacagaag
ctagtctccttattgatttgctgcgtacaaggatgaaaccgagtatactgtgctgccacaagtctgagcattgtcgagatggtggctgt
tgctgtcctgaggggttgggcaagctgaaaaagttcctgattgacttcttgagccatttgcacagagaattggttgggatgctaaaagcggg
gaggggtcacttggatgattttagaggtacactttaactgcttggcagaacttggccatgaagctacaataatgaagctgtacgccgat
ttaacatcttgggaagatagagaacaccactactccctccagatgttgcgaagggcgcatatgttgcttgatgcagacagtgaacaaatc
aaacagagccggctatgaatcactattgaagatctataaagaaactgatctgagccaggaaaagtcagaattttaggttcttagcatcttgc
cctgatcctgatgttctgtgatacactggacttcatgctgtcacctgaggtgaggaatcaggactctataattttagcttagaggagtcggtgca
gctgggacatgaggtggcatggacatggtgaaggaaaagtgggactacatttcagacaccttcttggaaacttcttcacctatttcttcca
ccaccgtctaccgctgcgcaccgatgaaatggcgacgacgggaggagtttcaagagcaggacgaaggccaacatcgcgaggagac
ggtgaagcagagcatcgagagagtgaggatcaatgccaaatgggtggagagcaccaggggccgaggctaacctgggcaatgtcctcaag
gaaatttctcacgaccactga
>22823 OsPN22823
atggacgagtcaggcagaggaacaggtgatgatcatggccgggagacgaaggatgctgcggcggcggcgtcgtcgtcgtcggggaag
aagggtgccgtgttcagctgttccggtacgccaccgctcgacgtgctgctgatggtgtcggcacgggtggcgcgctcggcaacggc
atctcgcagccctcatgacggctcttctggcaacgtcatcaactccttcggcggaacaccagcggcagcgctctccgcagcgtgacca
aggtgaaggtatcatgctggactatggcaggagaagggcagctctgccgcacccgttcttataacctgaaagccgttctgagcaggatatta
cattcttcgacacagagatgacaactggtgaagcagttctagaatgtctagtataccctcctaattcaaggtgctcttggtagaaggagg
gaagctgtagaactgttatcaagcttcatcggtggcttcatatagcattcactagaggatggcttctcactcttgcattgtaacatcgctacc
gttaattgctattgccagtgcagtttctgcacaggccctaactagagtttctagcaagagacaaacatcataatgatgctggtgggacacagt
gaacagaccattggatctataagaacagttgtgtccttcaatggtgagaagaaagcgatagcaatgtaccgtaatttataagaagtcataca
aggctactattgaggaaggcattatcactggtttggcatgggctctgcatgtgcgtctgatttggcagctatggattagccttctggtatggtg
gaaagctaatactaggaaaggttacacaggaggaaaaatcatgactatctgttggcgtgttgaccggcaatgcaacacctgcagttgctg
cagttgtggaaggtcaatctgcagcatacaatttgtcaaaacaattgagaggaaccagagatagattccgatatacaaatggcatggttt
agaagatatgaatggcgatattgagctaaaggatgtgtacttctgtacccgtgaagaccagagcagttgatattggatggattgtcgttaca
gtagcgagtggaacaacaatggctatagtcggagagagtggaagcggaaagtcactgttatcagcctagtcgaaagattctacgatccac
agtctggcgaagtttaatatagatggaattgcatcaagaaactgagacttgattggataagagggaagatcggtctttagccaagagcctc
tgcttttatggccctcattaaagataacataatataatggtaaaaaagatgcaacgcttgaagagatcaagagagcagcggagcttgcattg
cagctaacttcatgacaagttaccaaatggttatgatactttagttggccagcgcggtactcagctctctggaggacaaaaacagagaattgc
aattgcaagagccatctcacaagatccaaaaatcctttgctcgtatgaagcaacaagtgcacttgatgtggagtctgagaggatagttcagga
ggcactaaatagaatgatggtagaagaaccacactcgttgcgtcactggttgagcactgtgaggaatggtgattgcatcacagtcgtccg
caaaggaaaaatagttgaacaaggctcctcatgatgcactgggtgaaggatcccgatggagcttactcccagctaattaggctacaagagactc
atcgtgatgaaaggcataaactaccagattccagatcaaaaagtactagttgtcattcagacgatcaagaactaaagatttctcagcaagag
caacagggtattccttcaagagccccttaggattgcctgttgatatacatgaggatggaatgacaagcgaacaacaaaaggttgaccactctga
cagtaaggccattaaaaaacaccatttggacggcttttaacttaataagccagaagtgccagttctttttaggttctatagcagcatcagt
gcatggagtcatttgcactatacgggtataataatgccaggtgttctaaatcattctatgaaccgccagatcagctgcgaaaagattctagat
tttgggcattgatgtctgttcttgggggttcttgttatttcaatccagcagaatatttttgggaattgctgggggaaagcttatacagc
gtgtccgtacactgtcatttcaagaattatgcaccaagaggttgccttgggttgataagccctccaattccagtggggcacttggtaacaggctc
tcagtcgatgcgttgatgtccgccgtttagtaggagataacctggcccttatagtcagggtgtagctacactaatcactggcttggccatagc

Figure 15
Page 3 of 6

tttgcggcagattggaggcttgcactgatcatcacttgtaattccttagtgggtgcacaggctatgctcaagttaagttctgaagggggt
cagtgaagaatctaaggagatgtatgaggatgcaaaccaagttgcggctgacgtgtaggcagcatcagaactgtagcatcttctgttcag
agaaaagagtgggtggcaatatacaacaagaatgtgaagctttaagaaaacagggaattcgaagcgggaatcgttggagggttggcttaag
tttctcaacttgatgtatctgacttacggtcttctctatgttgggtgcaaagttcgttaagtcagggaactacttttcagatgtttcaaa
gtttcttctttagtttggcagccgttgggtttcgcagtcagtgcaagtgtcwtactaatgcaacaarggcaagggttctgccattccattt
tagtattatcgaicggaagtctaggattgattcaagtagcgacgaggggagcgataatggaaaacgtcactggcagcattgattcaataatgtc
agtttcaagtacccalcacgccctgatgttcaaataattcagtgactttaccttgcacattcctccaaaagaccatagcactgttggagaaagc
ggtagtggggaagtccacgataattgcttattagagcgtttctatgatcctgattctgtaatatctcactagatggagttgaaattagaagcttaa
aagtcagctggttggaggatcagatggggctttagggcaggagccagtgctttcaacgamacaatccgtgccaaacataacatacggga
aacacagtgaggtaacagagggaagagatcactgctgtggccaaggcagcaaacgccatgagttcgtatcaagcctgccacaaggatcac
gacacagtggttagtgagaagggttgcactatctggtgggcaaaaacagagggttagcaattgcaaggccatcctaaaggaccctaa
gatactgctacttgatgaggcaaccagtgctctagatgcagaatcagaacgtgttgttcaagatgattggatcagtcagtgatggtcaacaggac
taccattgtagtggcacaccgccttccacaatacaagggggtgatatgattgcagtcctcaaggaaggaaaaattgcagaaaaaggaaag
catgaagcactattgcatcaaggatggtgcatatgcttactgtacaactccgctctaattccgagtaa
>22825 OsPN22825

atggcggaggcggaggcggggagggaaggagaagaaggaggtggtgcggctggagcgcgaatcgggtgatccccatcatgaagcccaa
gctcatcatgaagctcgctacctcatcgaacaacaatctgacagagaagaattctgaagctctgcaaggaggattgagtacaccattagg
cctggtatcatctccagtttgatgatgatggaactatttgccttttgacctgttcatgggtgccaaaaactgcaacagcagaatttctgcac
tgaggaaagtcgatacctcagcagaacttctgacctatttcttcagggtgatgcaaaaaagcaatttcaacatattatcagatgatgaggtt
agcttgctcattctggacaatatctgctgaatctccatcaaagttgatgaagcaaaagtagatacaaaacttttgcgaagtattttaaaggaca
ccatcacgacaacctaccggagttctcagataagtatgcatattccggaggggcattggactggatcgactagcaacttttctttaggag
aaagtggacatgatcatagctcgtgcgtggcgtggttcttggagaaaacaagactgcaaaagctcttctcaagaaagaaaagtgttaggcc
aaagacagattccaagaagaatgatgatctagtt
>29041 OsPN29041

acaaactccctgggtgcaaatcagaagggcacttcgtgcccttggatgatcacacggatgggttgatggatttgaagatactgaagtgaga
tcttcggaggaaactgatgcttggagggaagctagattggcaatcagcaagtagtgcataccaaaagggtgaaagcgtgcagctactgccca
gaccaccaagcatcatcgcttccaggtggtatctagtcgaaagtttcaagctcaagtggaatctataggccaggagccaaacgcacgtcct
cagaattcttccgaattcgtaggcgtggaagaaggcggcaagctgtcaagcaagaagctgcaactgagctcacagattcagacaattca
gatgacatggac
>29076 OsPN29076

gagagagatatcagtatggttaacagttatccccatctgcttttgattgcaaagggttgaattcagcatctatttcatattatgaatctgatgagg
aatccagtgcccttctgactggttaagagactatgacccaagggaaggaaactgaaagactctatctgcagtggaagacatttctgtc
atcaaagcagttctcctctcgttgatcctccaatttctggtccaagggtgaacagcttatggagcttccaaatccaaggctgcagtaattt
gaacaaaagtatggtctgaattggaccaggacacaagtgatctaccgaaaaagcagggaagaagataaagttaagttctgaagataga
acatactgctgcgactgttggaaacctgtatggccttctcgtataccattgttcaacatgcatgagacttatctcatatctacagagtttgaagggc
ataatgatggaaaatgcagtaaaatccaccagtcctgatgaaagcagagaaaatgatgagccaaaagtgaagggtaccaaatctgatacg
aaggaaaaagattcccttgatgtagctctgtcattgaaccgtccagtgacagaaaattgatgcaatgccatagactttgaagaaatttga
gaaagttcgtcacaatgattcgaacaaggagacagtaaaagcagattggacttaattgggtccaatggagttccatcatttgccttcacctgc
atttttcttgaacctgcaattgtcaaaagtcaaaacmgaaaagatgatgaactcaaggattggacctcctctttagaggaaatgcaatgcaatg
tctgctcaaaagctagttcaagaggttccaaatctggtcacaagttgtcctggcaatgtgggtgatgagaaggtgcaaaaatctaaaaagccc
accctgataatacttctggtgagggaagcacattctacaacaggcaagccaacacgattgctagctgttaatgggggggttggttccagagtca
tcttgaggcctcttataggacgaaactctcacatacttaagcagcaaaaagataaacttgccttgacatcgaagcagccttgcctgaggaggca
ctgagagcctcaaaggtcagcaataagaaggcgttcattg
>29077 OsPN29077

gttgaagctcagaagaaaattgaagctgcaattcggcagaaaggtattgatgaaaactgggagggtgccttggagcacaacctgaagcat
ttgctcagtggttatgctgtatgtggacatggaagtcaatggtgtaccttgaaggccttgggtgacagtgaggcacaatcaactatcatatca

Figure 15
Page 4 of 6

aaaagctgtgccgaacgttgccgggtgcttaggctgctgatcagcgatataagaggtgtgctattggcgttggtcaatcagagattcttgta
gaatacatgtagcccaataaagattggccatgtcttctatcctgttctacacagattggatgtcccaatatggaattcttttgggcttgata
tgctgcggaaacatcagtgataattgacttaaaagacaatgttcttagagtaggaggtggtgaggtttcagtgcccttctacaagagaaaga
catccctcacatatacgtgatgaagagaaactatcgaagctagcatctctcagccaaggagctgctggagaatcgtcaacggcacgggag
aagactcctgatgcgccaccagagctcctactacaggtgctccagctgtgaatcctccgaaccacaggaggaggagattcgaagca
aaagtcacgaaactggtggaacttgattcgaccgggcatccgaatccaagctcgaagtgtcaacgggaa

>29084 OsPN29084

gaatctggaatcttcaggcaaatcttagaggcaaatctgatttgaatctgatccgtggccttagtatactgatagtgtctaaagatctagtcctg
aatatgcttatccgtgatcctacaaagagatttactgtcacgaggttctctgtcatccatggattgttgatgctgtggcacctgataagcct
attgattctgtgtttgtcaaggctgaacacttttctgcaatgaacaagctcaagaagatggctttgagggttattgctgaaagcctatctgaa
gaagagatcggagggttgaagggaattgtttaaagtatcgatactgacaatagtgggacgataacttatgatgaactgaagaatggcctgaaa
agggtgggtcagatctaatagaaccagaaatccaggctttaatggatgccgtgatattgacaacagtggaaccatcgactatggtgaatt
cttagcagctactttgacatgaataaactggagaggaggaaaacttggtgtcagcattcacgttcttgataaggatggagtggttcata
accattgatgagctctcgcaagcctgcgaacagtttggccttctgatgttcatcttgaggatgatcaaatgtggatcaaaataatgatgg
acaaattgtatcagcgagtttccgcgatgatgagaaaggcaatgctgggtggagccaatgctgggtgagtcacttctactggtggaaca
ggaagga

>29086 OsAAB53810

atgacgctggtgaagattggtccgtggggcggaatggagggtcagctcaggacatcagtggtccaccaagaagctgttaggcgtgaca
atctacagctcagatgcaatcagatccattgccttcaactacatcggtgtggatggacaggaatatgccattggtccatgggggtggggcgga
aagcacctctacagagattaaactgggctcctctgagcagatcaaggagatttctggaacccatggcccagctatgatctggtgacattgt
cacctatcttaagattgtgacaagtgtctaataacatcacagggtggtgagtcctcaaatggaagggaattcagcattccactgcaagactctgg
ccatgtcgttgattctttggaaggcttgaacgcttatcgacgcaattggcatctacgtccaccctga

>29098 OsPIP2a

atggggaaggacgaggtgatggagagcggcgccgcccggcgcaattcgccggccaaggactacacggacccgccggcgccgc
tgatcgacgcggcggaagctgggggtcgtggtcgctgtaccgcgccgtcatcgccagttcatcgccacgtgctgttctgtacatccgt
ggccacggtgatcgggtacaagcaccagacggacgcgtcgccctccggcgccgacgcggcggtcgggcgccgtggcggtgctcgga
tctcgtgggcttggcgccatgatcttcatcctggtctactgcaccgccggcatctccggcgggcacatcaacccggcggtgacgttcgg
gcttcttggcgccgaaggtgtccctggtccgcgccatcctctacatctgtggcgagtgccctggcgccatctgcccgtcgggctcgtc
aatgcttcccaaacgcttacttaacaggtacggggcgccgccaacaccctcgccggcggtactccaaggcgaccggcctcgccg
ccgagatcatcgccaccttctgctcgtctacaccgttctcggccaccgaccccaagcgcaacggcgccgactcacatgtccgggtctt
ggcgccgtgccaatcggttcgccgtgttcattgtccacctggcgacgatcccgatccggcgaccggcatcaacccggccaggagcat
cggagcggcggtatcttcaacaacgagaaggcgtggcacaaccattggtatcttgggtcgggccgttcgtcgccgcccatcgccg
gttctaccaccagtatcctcctcgggcgccgcatcaaggccctcggtccttcaggagcaacgcgtga

>29113 OsPN29113

tgcaaaagttcatatattgacgcaattcttgggacaaccgtcaaggttccactgttgatgggatggttgacctaaagatccctccggtaccc
agccaggcacaacgttagtgatgtccaagaagggtgtccactcctcgggaagtcaaatgctcgtggagaccagctagtgcgtgtacaggt
cgagattccaaagcgtctgagcagcgacgagaggaagctgattgaagaacttgcaaacctaacaagggtcaaacagcaaacagcagga
gataa

>29115 OsPN29115

gtctccttcccttactccccacggccggcggttggtgcaggcgctcgggctcgcgggttagccccgtcgtggtcgccggcgccgc
gggcaccagcggttatgggggtcgttgactaacaccagggggtcaggttggagtggttggcacggttcaatgagattgtaccaactt
gctactgcagggagctcttgagacattgagagatattccgtcaaaaaagaaaataacagttctaagtgttcctggcagtttgaattcctgt
tgcggcacaaaagcttgggaagcttgaaagtttgatgccaatttgtgcatggcgctgtgcattcagaggtgacacaaccactacggtgc
tgttgcaactctgctgctcaggtgtactgtctgtgattgtctgtgagatcccatgcatatttggtgttctgacatgtgatgacatggacca
ggctctaaactgtgtggtggaaggctggaaacaaggagctgaagccgctctaactgcaatcgagatggcctcgtgttccagcatcac
ctggcctga

Figure 15
Page 5 of 6

>29116 OsPN29116

ggaaggaagcctgttgcggcccaacattacaacaagggtgataaagatggaacaagcactggcagccgatgcacttctgttgcattgggta
ccggagcgtgaaggaatcttgtttagccattctgaiggaatctgtatgtgtacgataaatgcaaagatggaaacactgagtgtacattcc
cagctataaaggatccggctcagttaatgattcacatgcaaagtcagcaagagcaaccaattgctagatggcatgttggcaaggttcaa
tcaacgccatctctttaccagatggagcatatttagcaactgttgggcgagatggttattgagagtgttgactttcgaaggagcaactaa
tatttggggaaaagtattatggtgcactattgtgtgcacatggagctcggatggcaaatatctgttgactggtggtgaagatgacttgtgc
aggtatggagcatggatgacagaaagatagtcgcgtgggg

>29117 OsPN29117

atgctggtgcagcgcaggagcggcgacacgggtccggccgtcagggtcagggtctccacggcgccctctccgcgacgtcggcgtgc
cggcgccactccaccttcggtgaattgaagggggtccttaccaggaactggcgtagagcctgaaaggcagaggctcttctccgtggga
aggagaagagtgaacatgattcctgcatacagctggggtcaaggatggagcaaaacttcttacttgagaagcctgccctgccaatgta
gagcagaggccgagccagtaattatggatgagagcatgatgaaggctgtgaggctgttggccgtgtaagagctgaagttgacagactct
ctgccaaggtatgtgattggagaagagtgttgcaggggagaaagattgaggataaagatttgttgccttgacggagcttctatgatggag
ctgctgaaacttgatggcatagaggcagaggggagaagcaagggcacaaaggaggtgaggtacgccgtgtccaaggctctgtggaga
cgttggataagctgaaggcaagaaatgccaatcccttcagcgatcaaaacaaatctgttcagtgaacgcagtgaggagacgttcgacaat
ggcatgggcagcttgaatgcacccccaccacgggttcttccacacaaataaacaccgactgggagcaattcgactag

>29118 OsPN29118

acctctcaggagatcagcagatggtgacgggtggcggagaggttcccgcgggaggtgagctcggagcggtgttccggtgcgtgcggc
tggggccggtcgaccaggccgagggcggaggtcgcgtaccagacggccgtcagcatcggcgccacgtgtcaagggcacacctgcacg
acgtcggccccgaagccctggcggtcgcggcgggcgccgagcagtgatccacttccgcctaccggcgacgggtcctcgcgga
gcaccggcgcgccggcgagggcaggtcgcggcgggcgccgcaacatcatcgtgtatcgggtgtggtgatggaccgtacccgacgc
ccggggccctacggcgcttccccgccggcacgccattctccacggccacccgcggccgtga

>29119 OsPN29119

atggagttcaggcggacggcgcccggtggccggagccgcgagggcagcgtgccggggcgccgcccgtggagcgtggcgatcgcgc
cttcccctcgcttcgattcgtctcgtgccctaagattgctgagagagcttggctcaaataaacagaagatttgggtgtgctgatgcaaatgtc
tatcatttctgaaacatgatgcctgtggttgttaataatccatagctagtgggacaaattatttgcgtgtattggaagagatgacacttca
gattaataagtgtggaagggtgatgcttggcttgaagagatgtgggcttgagcgaagcagtttaagatgcagtcataatctgatccatga
gtctgtcccagttgctactaaactgttgcgtgtaaaagtattgaaacatggatcctgtgttcgcacctaataccaaaagtaccggatgcagc
caactgaaggaaggaacaggaggtgttgcgtgattcaagattatctcaattccatccagccctaaccctgctgtctcgaagctgatgcaa
cagggtctcactccttcttggatatactgcagtcagcttgcgtcatcaaggatccttctagttggcaccattaattcttgcagctattgca
aagaatagacctgttactacgaacgcatactgccagtgctacttgggttaccctagtgttgagggttgcgaaggagctcatcctgcaagct
gcggtattccctgaaaacagctttttagggttctaaggagtccttgcaggcaatgatgagtcgaaggatactctggttaaggcagcttctg
ttctgagccctggtgaagcaacagaacagattattaggcaagtagagaagatgactagaaatattgagcgcgttctcgtgctagcaaggat
gaacctcaacgttggatatgccttacggagatgttagtcgga

Figure 15
Page 6 of 6

Figure 16 (Example 9)

>20215 Os008938-3209

atgtcgcctgctgaggcatcgcgtgaggagaatgtgtacatggcaagcttgccgagcaggctgagcgttacgaggaaatgtcgaattca
 tggagaaggtggcaagaccactgatgttggtgagctcaccgttgaggagcggaacctgcttctgtggttacaagaacgtgattggtgct
 cggagggcatcgtggaggatcatctcttattgagcagaaggaggagagccgtgggaatgaggcatatgttgcataaattaaggagtagc
 gtagcaggattgaaactgagctcagcaagatctgtgatgtatccttaagcttctgattccaccttgcccatctgccactgctgcagagtc
 caaggtgttctacctgaaaatgaagggtgactaccacaggtaccttgcgtgagtttaagtcaggagctgagaggaaggaagcagctgagaac
 actcttgtggcatacaagtctgccaggatattgcactcgtgacctgctacaactcaccgataaggcttgacttgactgaacttctcag
 tgttctactatgatactgaactcaccagaccgtgcttgcacacctgcaagcaggcgttcgacgatgctattgctgaactggacactcttgg
 cgaggagtcttacaaggacagcaccttgatcatgcaacttctctgtgacaatctgactctctggacctctgacaatgcggaggatggtggtga
 cgagatcaaggaagcagcgaagcctgaaggagaggccactaa

>20254 OsPP2A-2

atgagcagccccatggcggcctcgacgaccagatcagcgcctcatgcagtgcgaagccccccccgagcccagggtcagagcactttg
 cgagaaggcaaaagagatattgatggaggagagcaacgttcaacctgtaaagagtcctgttacaatatgtggtgatattcatgggcagtttca
 tgacctgcagaactgttccgaatcgggtggaagtgccagatacaactactgtttatgggagattacgtggatcgtggttattattctgttga
 aactgtcacgcttttgggtggtttaaagggttgcgtatcctcagcgaattactattctcagaggaaaccagaaagtcgacagatcactcaagttt
 atggattctatgacgagtgcctaaggaagtacgggaatgcaaatgtgtggaaaactttacagatctcttcgattactccccctgacagcattg
 gttgagtcagaaatatttgcctgcattggtgattatcgccatccattgagacacttgataacatacgttaactcagtcgttccaagaagtcccc
 catgaaggggcccatgtgtgatcttctgtggtctgatccagacgatcgatgtggttggggtatttctctcaggtgctggatacaccttcgggc
 aggatatacagagcagttcaaccataccaataatttaagacttattgctagagctcaccagttggtcatggagggttcaattgggctcatga
 gcaaaaagtgttaccatatttagtcacctaattattgctatcgtgtgggaacatggcatcaatcttgaagttgatgattgcagggagcata
 cattcatccagtttgagccagccccagaaggggagagccagatgtaactcgtagaacacctgactatttctgtga

>20311 OsCAA90866

atggctcaggtgtggaggaagtcagcaacaagatgcaggtgcagatgcgcctccaccccgccgcccggcgaggaggaggagcgcga
 cctccccgctccccgcccctcttcgacaaggcgtccacctccactccctcgctccagctcctccctcgaccaggaagggtatccgcaaggg
 ggtcgacctgctgcggcggtgcgacgagatggtcagtaaggtcggcctcttctcctccaacgagaccaaggatgacatctccaccgcaa
 cctcaaatacctactctgactactaccttggggaatgactgaagggttagcgcaggaagatcggtatccagctccttaaggcgtcgcag
 gatcatttgaaggaattcatttctatctgtgaagcactggagcttatacagaggatgagcttgaataatctaggcagaagaacctgatacca
 tggcaaatcgaagagcacagaagggtgcacggttcaagcgccaaaaggctgcagaacaagctctagaatcaaggagaggaaagaaa
 ggctgtaggcgtcatgagagcagctgcttcatcagctcctattgagctgggaggagatctttgaggatgatggagaaggaaagagaaggcat
 ggttagctactatctcattggctctatcgaaggcttgcacctcttgacatgttaaagaaggaaagaagaatcgttccggcagtaaaagaagga
 aagcgaaggatggtaatgcatttgctcgtgaaatgcttgatgaacgtacaaaagggtgaagcatggcaccataatgctgccaaccgtgc
 accatactccaaaccagctgatccaatcacttgcaactaacagtgaaagagagagaatggcagcacaagttttcagccaagttataggt
 tggcaacgatgagcatagaagaagctggcttacgtgagatgaaatgatggagaatggcaagaagaactgccaagatgattcaagaat
 caaactccgcatggcacaaggatggcagtcggtcagcccaaggagcagggacgccgaagaggagaagcagagaggatgggatgac
 tggaaaggacgataaccctcgtggtgcaggcaacaagaagctcactccctgtggctaa

>20618 OS011994-D16

atgttcggcgcgccgcaagaagagcgacaacaccaagtactacgagatcctgggggtccccaagaccgcctcccaggacgacctca
 agaaggcgtaccgcaaggccgcatcaagaaccaccccgacaaggcgccgaccccgagaagttcaaggagcttcacaagcttatga
 ggtattgagtacccggagaaacgtgaaatctatgaccaatatggtgaagatgccctcaaggaaggaatgggtggaggcggtatcccatgtt
 gatccatttgacatctttcatcattctttggacctcttttgggtggtggcagcagcaggggagaaaggcaaggaggggagaggatgtg
 atccatccgcttaagggttctctagaagatctttacaatggtacttcaagaagctctcttcttcccgcaatgtcctctgcgccaagtgcaggggc
 aagggttccaagtctggtgcttccatgaggtgccaggttgcagggggtctggcatgaaaatcaccatccgcccagctggggccttccatgat
 acagcagatgcagcagccttgcaatgagtgtaaggggactggagagagcattaatgagaaggatcgtgccagggtgcaggggcgag
 aaggttattcaggagaagaaggttctggagggtcacgttgagaaggggatgcaacacaatcagaagatcatttccctggtgaagctgatga

Figure 16
Page 1 of 5

ggcgctgataccgttacgggagacattgtatctgctccagcagaaggaccactccaagttcaaaaggaaggcgatgatctctttatga
gcacaccttatctctgactgaagcactttgtggttccaattgtctgacacatctggacaacagacagctgctcattaagtcaaaccccggtg
aagttgttaagcctgaccaattcaaggcaataaacgatgagggaaatgccaatgtaccagaggcccttcatgaagggaagctctacattcatt
tcacgggtggagttccctgattccctggcgctgaacaatgcaaggctctcgaggctgtgctccaccgaagcctgcatcccagctgacaga
aatggagatagatgaatcgaggagaccacgatgcacgatgtcaacaacattgaggaagagatgcgcaggaaagcccaagctgctcag
gaggcgatgatgaggacgatgagatgcctggagggtgccagagaggttcagtcgcgcaacagtaa
>21639 OsORF020300-223

atggtcgagggtggaggaagtcagcaacaagatgcaggcgagatgcgctccaccccgcgcgggcgggaggaggaggacgccc
acctcccgctccctgcctcttcgacaaggcgctcccactccactccctgcctccagctcctccctcgcacaggaagggaatccgcaaggg
ggtcgacctgctggcggtgacgacgagatggtcagtaagctcgccctcttctcccaacgagaccaaggatgacgtctccactgccaa
cctcaaatcctactctgtccgtactaccttgggaaatgactgaagggttagcgcaggaagatcgatcccagctcctaaggcgtcgcag
gatcattgaaggaaattcattctatctgtgaagcactggagcttatcagaggatgagcttgattgtctaggcagaagcaacctgataccat
ggcaaatcgaagagcacagaaggctgcacggttcaagcgccaaaaggctgcagaacaaagctcctagaatcaaggagaggaaaga
aaggcgtagggcggtcattgagagcagctgctttatcagctcctattgaagctggggagggaagatgctttgaggatgatggagaggaagaa
agagaggcatggttagctactatctcattggctctatgcaaggctttgaccttctgacatgttaagaaggaagaagaatgcttctggcagt
aaaagaaggaagcgaaggatggaatgcatttgctgtgaatgcttgatgaacgtacaaaaaaggctgaagcatggcaccataatgct
gccaacctgacaccatattccaagccagctgatccaatcacttgtgcaacatttgctcaagatgctattgaaggtagagcaagtgttcacaag
cacatgagcacaaacaccagccgctgatatttggccctgcaagtcttgttggtggaggactaaccagtgaagagagagaatggcagcac
aagttttcagccaagttacaggttgccaacgatgagcatagaagaagctggcttactgagatgaaaatgatggagaaatggcaagaaag
aactgctaagatgattcaagaatcaaatccgcatggcacaaggatggcagtcggtcagcccaagaggatgaggacgctgaagaggaga
aagcaagagcatgggatgactggaaggacgataacctctgtggtgcaggcaacaagaagctcactccctgtggctag
>23045 OsPN23045

atggcgccatattcttgccttctcgcgccactgcgcccggcgccgactgcaggccgtcgacggcgggcgggcgggcgggggcg
ggggcgggggcggtcgtgctcagcgtgaggccccggcggggggtcgcgctcggtggtgcgctgcgtcgccacggcgggcgatgtcca
cccactgtcgagaaacaaagatgaattttctcaagtcatacaagcgtcctatcctaagcatttacagtagcttacaagaactttgttaca
gcaacatctgatgatacaaaacaacatatcaatatgatcggtgttcttctgttggttgtgaccgtctatgaccaactcatggaagggtacc
ctagcaatgaggataggatgcaatcttcaaagcatatataacagcggttaaacgaagaccctgagcaatacagagctgacgcacaaaagat
ggaagagtgggctcgttccagaatggaattctttagttgagttttatccaaagatggagaaatagaggccattctgaaagatatttcagaa
agggcccagggtgaagggaagcttcagctacagccggttcttgcagttggcttgttccgttgccttgagcttgcaaatgcaacagagccaacc
atactagacaagctttgcgtgcttaaacatcaacaaaagaagtgtcgacagggtatcgcgatgtttaccggaacatactctccaaattggtcc
aggctaaggaacttctcaaggaatacgtggaagggaagaaagaagagagaggaaagatcagagaccccaaaatcgaatgaagctgtt
acgaaattgacgggagctcaattccatgaggcattaa
>23186 OsAAG46136

atgggttctgaaggaccttctggtgttaccgttcacgttactggattcaagaagttccatggagtcgctgagaatccgacggagaagattgtgc
gcaatcttgatcatttatggaaaagagagggttgccataaagggttaacacttggaagttgcactgttcttgagactgctgggcagggtgggct
tggtccgttgatgaagtgttgaatcagccatcgtagacaaagagtatgggtgaatgatcaggggcaagtattctgctcattttggagtca
acagtggcacaacaagggttgccttgagaatcaagctattaatgaagctaccttccggttgcctgatgagctgggatggaaccacagagg
gcccttatgtgtcatctgatggaagcatctcaaatgaagaaagaccactgtgcctgtgaatgaagtgaacaagtccttacaacagatgggc
ttcgaatggcgcccttcggatgacgctggtcgattcgtatgcaactatgtctattaccaatctttaggttcgcagaacagcgcggtatcaagtc
ttgttcgtccatttcccccttccacgacgattagtgaggaagttcagatgaacttcgtcgcaacctcctcgaagttcttgcctccagaactat
gcacagtaa
>23225 OsPN23225

atggagaaggatcaccagcccgtcatcagcttgcgccccggcgggcgggcgggcgggcgccccggcgccctcttctccccgcct
tcgcccggcgccctccggctccggcgactgctccgctccacgtcgcgggcgcatccaagattggtgatccaaattttgaggtgcgcga
gcgtgttcgtatacaagagatcaacttctcagctgcgtgagatcggtgacatacctgaggccatttaattaagcaagagattgatagag
cttcatggtgaagatcagatttgggggtgccttgagtcagatgtgcaagtgcagacacaaaccaggcgcaaccgcacaatcggtatggtg

Figure 16
Page 2 of 5

agacagataaccgtgactggcgtgcacgcacgggtcaacctccggcagctaataagagaagtcctgggacaacattcgtgaagctaaag
cagccatgcttcaagtggcggcaacaagagcaagtaacaggcaggatcaattaaaccaccagtttgcttcaaaagctcaggttggccc
gactcctgcttatacaaggctgaggtaccttggcagctagaagaggcaatctctcagagaaagatagagtcctgaaaacagtgaaggta
tactaaacaaactgacaccgggagaatttgatctactcaagggccaaactaatggaatctggaattaccactgctgatattctaaaggatgttatt
tctctcatattgagaaggctgttttgagcccaccttctgtccgatgtatgctcaacttgttccgatcctaatgagaagctaccatcattcccttc
ggaagaaccaggtggcaagagattacgttcaagcgcgtgctgtgaacaattgtcagggaagcattgaaaggcgtgagagcctaagggc
tgaaatagcaaaattgactggccctgaccaagagatggagagaagggacaaagaaaggattgtcaactaagaacacttggaatatccg
tctaattggtgagctacttaagcaaaagatggtacctgaaaagatagttcatcattgttcagggaactctgggggtctggacctgataaaaag
gcttgcctgaagaggaaaacgttgaggctattgtcagttcttcaatacaattggtaaacagcttgacgagaacccaaaatctcgtcgaatta
atgatacctacttaticagatgaaggagttacaacaacaccttcagttggcaccacgtctgaggttatgttctgctgatgtggtgatcttctgtc
aaacaattgggtgacctgacgtgaagagattaaggccaagacaatcagtgaaattcatgatgaggtataaagacacttgactacggcca
ggggcaacaggactcacgaggaatggcgtaatgcaccaggtggcctcttcccttgggtgattcccaatgaatcaccaggaacagga
gggatgatgcttgggatgacctggaacacctggtatgcctggatcaagaaagatgcttggatgacctgggttagataatgataactgggaag
ttccacgttcaaatcaatgccaagaggcgactcttctgtaaccagggtccgttgccttaataagccatcatctattaacaagccatcatctatta
actccaggcttcttctcatggaagtggagcactaattggaagatgctcttgggcagtggtggccaccatctcgccttccagccttatg
gctagtcaccactatcacaccagctcaaaactgcaccatcacaaaacctgtaagtgcagctccagctgttgcctgtcacagacaaagcagc
cggttcttccatgagatgccagctgcagttcagaagaagacggatcccttcttgaagagtatttggcatacgtattctggtgaagcaca
caatgtatcaggaggttcagtgctctgagtactactctgagattgtgaaagaagctatcaatcttgccttgataagggtcccaacttcattga
tccccgtttaggtgctggagcatctgcacaccaagaaaatttcaagactgaggacctgaaaactggatgcttactgtatcgaccctgctg
gaagatattggtattgaccttctctagctccagccttgttgggtgaagtgttgacgactgagttgtcgtgtagcttgagcttgaagttgtga
ggagattctaaaagcagtgaggagatacacttccgcaaaaggaaattttgatgctgtcatgaaaacctgggtggaaactcttcaggtcaggc
tatcttgagctcgcagctgtgtgtaatcgacgcctgcaacaaacttctgaaataa

>23266 OsAAK63900

atggcaggtgctcctcagaggactagtctcctcggcggttgtgccgtcttgatggcggtcggcgtcggcgaggaggcggtcgggtggtc
tcggcacggccaagtgcgccgactgcacccggaagaacatgaaagctgaggatgcttcaagaacctccaggtggcgalttaagtgcag
aacggcaacggcgagtagagagcaaggccactgaaagctcgacggcaccggtgccttcagcgtcccccttgacggcgacctccaca
gctccgactgcatcgtcagctccacagcgccaccaacgagccatgccccggcaggagccatccaagatcgtgccaatgtcggagggg
caccttctgctgccgtcgccggcaagaccactaccgatcggcggttgcgcgtccgtgacctctgcggccaatcaagaagaagatcata
gaccacttccacaagaagccggtgccaccaagccggagccaaagccggagccaccaagcccaagcctgagccggagcaccattc
ctcgaccacatccacaagaaggagaagcacttctcgaccacttccacaagaaggccgtgccaccaagcctgagcctaagcccgagcc
caagccgagcctgcgccagagtaccacaacctgacccctccggcgcaattaa

>23268 OsPN23268

gtgctggagacgctgatcggcggtgaccacttctcggaggaggaggcgaggcgacgctgcggctgctgctggaggaggagaacgag
gcgcgcacgcggccttctcgtgctcctcagggccaaggcgagacctacgaagagatcgtgggttggcgaaggcgatgataggctg
ctgctgccgggtcgacgggctcgtgacgccgtcgacattgtcggcaccggcggggacggcgcgacaccgtcaacatctccaccggc
tccaccatcctcggcgccggcggtgccaagtcgccaagcaaggagcagagctagctcgtcgccctgcggcagcgccgatggta
ttaaagatgtgtcaatgaggtgggtgttgattcatgatgtctgcaaaactatcaccagcaatgaaaatcgtgaaacctgtgaggaagaagct
gaaaataaagacagtttcaatatccttggctccttcttaaatccagcaagggtgccttatgctgttattggtgtttaccacgagaacatagtacc
aagatggctaaggcagctcagaaatttgaatgaagagagcattggttgcattcgaaggggttgatgaataagccacttgggcctgg
atataccttgatgtcacaccaagaaagattgaaaaatgctcttcgatccattggtattgtatccccgctgcacgctggaagatctaaaag
gaggtgatccagcgttcaacgcaaaagtctccaggatgttctgctggtaagaggttcaattgctgatgctctcgtgctgaatgctgcagc
atctcttctgtcagtggttaaagtaatacttgcagatggtgtagcgttagcacaggaacacaacgctccggggaggccatcaacacatt
gaatcctggataaagatttcaatgtaagtacttctgataattaa

>24162 OsPN24162

atggccatgaagggacccggcctcttctccgacattggcaagaggggccaaggatctgctcaccaaggactacacctatgaccagaagctg
accgtctccaccgtcagctctccggagtgggcctcacttccacagctgtgaagaaggggggttatacttcttgatgtcagctcagtttac

Figure 16
Page 3 of 5

aagtacaagagtactctcgtcgatgtcaaaaggacacagaatctaatactctactactttgactgtgttgatgtccttccatccacaaagctt
 gtgacaagtgtaagctgcctgactacaattctggaaaggtggagatgcaatacttccatgagaatgcaagtttgccactgctgttggcatga
 agccatctcctgtggttgaatttctggaacagctggctgctcaaggacttgccttgggtgcagaagctgggttgacactgctacaggaaaattt
 accaagtacagtgtcgcgataaggtgaaccaagccagattatcatgctgcaatcgttctggcggacaaaggtgacaccgttaaagtgtcagg
 tgtatatcaccttgatgataagcagaatcctctgttgttgcgaattaacccgcaggctcicaacaaatgagaacactctactgttgggtggcc
 tgtacaaggttgaccccagacagctgtgaaggcaaggctcaacaacactggaaagcttgcctgcttctccagcatgaggttaaacccaa
 gtcggtttgacaatctctggtgaattcgacaccaaggccttggacagaccccccaagtttggcttagcgcttgactcaggccctga
 >24775 OsCAA33838

atggcgagttccgttttctcctggttttctatatactttgtgttcttattatgccatggttctatggcccagctattaatcccagcacaaacccat
 ggcatagtctcggcaaggaagtttagggagtgtagattgatagactacaagcatttgaaccacttcggaaagtgggtcagaagctggg
 gtgactgagtactcgtatgagaagaatgaattattccagtgcacgggtactttgtgatccgacgtgtcattcagcctcaaggccttttggctac
 cgatacacaatatctcgtggttctacatcatcaaggagaggttctatgggtttaaacttccccggttgccttgcgacttaccagcaac
 aattccaacaatttctatctcaaggccaaagtcagagccaaaagtttagagatgagcaccaaaagattcatcaatttaggcaaggagacattg
 ttgcactcccagctggtgttgcacattggttctacaatgatggtgatgcacctattgttgccttatatgtttatgacgtaaacacaacgccaatc
 agcttgaaccttaggcaaaaggagtcttattagccggcaacaacaatcgggctcaacaacaacagtatatggtagctcaattgagcaaca
 ctctgggcaaaacatactcagcggatttgggttgagatgctaagtgaggcttaggcatcaacgcagtagcagcaaaagggtctacagagc
 caaaatgatcaaaaggagagatcatacatgtgaagaatggcctcaattgttgaaccgactttagacacaacagcaagaacaagcacaag
 cacaagatcaatacaaaagttcaatacagtgaaacgacagcaaacatcttctcgatggaacggattggaggagaactttgacgatcaag
 gtgagagtaaacattgaaaatcctagctgtgctgattcatacaaccacgtgccggaaggataacaagtgtaatgtcagaagttccccatc
 cttaacctcatcaaatgagcgtaccagagtaaacctataaccagaatgctattctcgccttctggaacgtcaatgctcatggttggctat
 atgattcaagggcgaictcgagttcaagtcgtagtaactttgaaagactgtgttgatggtgtccttcgccagcacaattatgatattccg
 caacattatgctgttgaagaaagcagagcgtgaaggatgccaatatcgaatcaagacaaacgctaacgccttcgtagccaccttgc
 agggaaaaactcagttccgtgccttgcagttgatgtatgcgtaatcgctatcgcatctcaaggagcaagcccgaagcctcaagaaca
 acagggggagaagagcacgggtgccttctactcctagatttcaacaacaataactaccaggattatcgaatgagtcgaaagcgagacctcaga
 gtaa

>26645 OsPN26645

atggcgacgcggctgctgtgttgacggcgctcctcctccccatcatcgccgccaccgcccgccctcgccttcccagggcgtgccc
 gtgcctactcggcgaggaaatccttgggcctggcgccacatgcacgacgtggatcgccgcgccgaccctgtcggcgtcatcgagg
 gggatgaggttaacattggcgaaggccattactcttccatgaacaaagatgattacattgctgtactcttctatgcttctggtgccattct
 cacaagaatgcaagccaaattttagatattagcatcttggttccatctattcggcatttgcgttgaagaatcctcaatcaggccaaagtataat
 atcaagatatgggattcatggtttccaacactatttcttgaattcaactatgcgagtgcggtatcatggaccacggactgttaagtccttggc
 tgccttaccgtgatgtttcaggttctgatgtttcaatgacgtcagaggccgtgctacattcagtagatggtattgagcttaagaaggatgctga
 acagggaaaattgtcattctggtgggcacgtcaccggagaaaatacttcagcaggatacttatctagcactggcaactgcatttgaatcttgc
 aggttactgtatcttcttccaaagataggttccttcgcaaacggcgctggaggaggcatactcttttccaaacttgggtgggtgtgcatga
 atatttctcacctacctgaacaagcaagacacaaattcttaggttatacccttcgaagcgagggaatttacaggaaggggccaggaatgc
 cactgcttgggcttcaagtcattagcatccgtctcaatcggagaaccaagcactattggaaggacaaactctacaatatgagctaagataa
 >29883 OsPN29883

ccacgcgtccgtgtcgcctctgctcctcctcctcgtccatctcctcctcatatcccactcgcgaatctcgattccctcctccgctctacct
 cgccggagttcaccttgttgggactgggagcggcgggcggtcggaggcaatggacgcttcgtcgggtggcgtctacggccagctca
 aggtgctcaaccattcttctttagctgggcctaatgtgattgaatcagaggagcatgctctgaagatggcaaaacacatcaaggcatca
 caaccaagcttggctgccacttgtgttaagtccagctttagataaagcaaatcgtaacatcaaaaatccttccgtggctcctggtctggaggaa
 ggctgaagatacttgaaaaagtgaagcaacatatgacattccagtgtgctactgatgtcatgaaagccaccagtgtgaagccgccggaa
 gaggggccgacatcatacaaatccagcttcttctgctgccagactgatcttctagtggtgctgccaagactggaaaaattatcaacatcaa
 gaaaggacaattctgtgctccttctgttatggccaactctcgagagaaaatcagacttgcctggaaatcaaaatgttatgctctgtgagcgtgg
 caccatgttggctacatgatctaattgtgatccaaggaattttagtggtgagagaagctaattgtccagttgtagctgatgaacgcatgc
 tctacaacagcctgctggaaaaagcttgatggtggaggtgttgaagtgggggcttacgagaactaataccatgcatcgcaaggacttctg

Figure 16
 Page 4 of 5

ttgctgtcggagttgatggtatttcatggaggtacatgatgatccctgaacgcaccttgatggcccaactcaatggccactgcgcaattg
gaggagctattagaggagctgattgcaattgctcgagtcaccaagggaagaagccactcaagatcgatctacccccctcaaagaat

This Page is inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ BLACK BORDERS
- ☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☐ BLURED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☒ COLORED OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REPERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images
problems checked, please do not report the
problems to the IFW Image Problem Mailbox**